PROBING THE DYNAMICS OF RNA POLYMERASE AND HU IN LIVE *E. COLI* CELLS

by
Kelsey Elaine Bettridge

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
April 2019

Abstract

Diffusion is at the heart of every biochemical process. Millions of proteins must navigate the heterogeneous, crowded cellular milieu to perform their various tasks. This molecular crowding has a significant effect on the diffusive behavior and kinetic rates of proteins and biochemical reactions. A powerful technique to understand biochemical processes within the context of this heterogeneous environment is single particle tracking (SPT). In Chapter 2, I use SPT to elucidate the dynamics of the RNAP search process and transcription cycle in live *E. coli* cells. Using FRAP, I find that transcription follows a simple initiation-elongation-termination cycle with kinetic rates that closely match those in the literature. Using SPT, I probed the search process of RNAP and found three diffusive states corresponding to DNA-bound, diffusion within the dense nucleoid, and diffusion within the cytoplasm. RNAP exhibited confinement in each state and displayed a preference for a DNA-bound state, suggesting a grid search strategy. Additionally, RNAP displayed kinetics that were not consistent with steady state kinetics. In Chapter 3, I use SPT to probe the molecular mechanism of HU-mediated chromosome organization. Using genetic mutations that abolish the various binding modes of HU, I find that HUαα and HUαβ displayed differential dynamics. Additionally, HUαα seems primarily responsible for non-specific binding while HUαβ seems primarily responsible for repressor loop formation. The kinetics of HU were highly transient, indicative of their non-specific binding across the nucleoid, and suggested a mechanism by which cumulative forces of thousands of HU are able to achieve chromosomal organization, a marked departure from the long-lived binding of other DNA organization proteins such as histones.

Primary reader: Dr. Jie Xiao

Secondary reader: Dr. Bin Wu

Acknowledgements

I would foremost like to thank my advisor, Dr. Jie Xiao, for providing me the space and resources for my two projects, which both took circuitous routes to publication. Even so, Jie continually provided insightful feedback, relentless optimism, and an infectious enthusiasm for science. In addition, I am extraordinarily grateful for the lab she's cultivated; everyone is inviting, kind, and generous. Xinxing, thank you for indulging me with conversation and my incessant questions every day. Chris, thank you for being my "lab husband"; our science philosophies are very similar and I'll miss sitting next to you at seminars. Gina, thank you for your constant emotional support and wisdom, and for helping me troubleshoot biochemical assays. Jason, thank you for your continual kindness and your ability to calm everyone's nerves. Xiaoli, thank you for your uncanny ability to make sense out of extremely complicated datasets (and complicated issues to boot), your insight has been invaluable for my projects. Ryan, thank you for being the social glue of our lab, and for our regular car rides where we can just chat about whatever. Josh, thanks for being the biggest lab cheerleader, it makes presentations easier. Drew, thanks for the coffee chats and our shared enthusiasm for delicious baked goods. Thanks to previous members- Jackson, Carla, Renee, and Max- for their help and support, I miss you all.

I am grateful for my thesis committee- Dr. Wolberger, Dr. Green, Dr. Ha, and new additions Dr. Wu and Dr. Wu- for their advice on both my work and my career trajectory. Casey, Tammy, Cathy, and Nicole deserve extensive thanks for handling administrative logistics; their behind-the-scenes work is the foundation

of Hopkins. I would also like to thank Dr. Espenshade and Dr. Ziegelstein, who have both given me invaluable advice on handling difficult situations and helped shape my leadership skills. In addition, I would like to thanks Dr. Greider for taking a risk on a biology novice and solidifying my desire to pursue my degree, and Dr. Chris Viggiani for patiently explaining essentially all of biology to me.

Finally, I would like to thank my friends and family. To my BCMB family- Risa, Mark, Laurel, Anthony, Katie, Aditya, Catherine, Jarrett, Stefanie, and Tom- thanks for “adopting” me into your fold, I treasure our friendship and miss our weekly lunches. Lauren, Marie, Anthony V. and Karole deserve my gratitude for hallway and experimental support. To my parents, Jeff and Mary, and my sister, Megan- thank you for your unconditional love and for always believing in me even when I didn’t, I love you with all my heart. Last, but definitely not least, I would like to thank my husband, John; without him, I wouldn’t be at Hopkins. Thank you for pushing me to be my best self and giving me the courage to go out of my comfort zone. Without your love, support, and constant goofs, my journey would have been infinitely more difficult. I am the person I am today because of you. I love you.

Table of Contents

Table of Tables

Table of Figures

Chapter 1: Introduction

I. Transcription cycle

The central dogma states that the flow of genetic information is linear: first, DNA is transcribed to make RNA through a process called transcription[1], which is then recognized by a large riboprotein complex known as the ribosome and used to synthesize new polypeptides, a process called translation[2]. In mammalian cells, newly synthesized RNAs are exported from the nucleus to the cytoplasm, where they are subject to numerous post-transcriptional modifications, degradation, or translation. However, because *E. coli* cells lack membrane-enclosed organelles and nascent RNAs can immediately be translated into protein, much of the gene regulation mechanisms focus on transcription. In *E. coli*, a single RNA polymerase is responsible for all transcription. The core enzyme, which consists of two alpha subunits, one beta and beta' subunit, and one omega subunit, has poor recognition of promoter sequences[3]. In order to confer promoter recognition, the core enzyme must bind a transcription factor known as a sigma factor. There are seven sigma factors in *E. coli* that recognize different subsets of promoters based on the environmental conditions. In fast growing media, most transcription comes from housekeeping genes[4], which are recognized by the σ70 transcription factor. The core RNAP complex bound to a sigma factor forms the holoenzyme. As the copy number of sigma factors greatly outnumbers that of RNAP[5], it is assumed that the majority of the pool of RNAP exist as holoenzymes.

Once the holoenzyme is bound to the promoter- known as the "closed complex"- it must unwind the DNA at the start site of transcription, a 13bp segment called the transcription bubble[6]. This unwound DNA-holoenzyme complex is known as the open complex. Once the DNA is unwound, RNAP can begin to synthesize RNA using the template as a guide. RNAP can synthesize around 8-10bp while still remaining in contact with the promoter sequence. The stabilizing forces of RNAP-promoter contacts can prevent RNAP from synthesizing past this point, and these short RNAs can be released from RNAP in a process known as abortive initiation[7]. The RNAP can repeat the initiation process until it can synthesize past 8-10bp and become removed from the promoter sequence, called promoter escape[6,7]. After promoter escape, the holoenzyme continues to synthesize RNA. The sigma factor bound to the core RNAP stochastically unbinds from the holoenzyme[8], leaving the core enzyme to continue transcription elongation. During this process, accessory factors, such as GreA, associate with the core RNAP to ensure elongation efficiency[9].

Finally, transcription can terminate through two processes: Rho-independent and Rho-dependent termination[10]. In Rho-independent termination, a GC rich hairpin forms in the nascent RNA, followed by a run of uridine bases. The weak forces of the dA-dU base pairs allows the nascent RNA to be removed from the RNAP active site, ending transcription. In Rho-dependent termination, the nascent RNA contains a recognition sequence for the ATP-dependent helicase Rho, usually a stretch of ~70bp that is unstructured. Downstream of this site contains a Rho termination pause site, stalling RNAP on the DNA to allow Rho to translocate down the nascent RNA and unwind the RNA-DNA hybrid to terminate transcription. About half of all *E. coli* operons terminate through Rho-dependent termination. However, other transcription termination factors have been identified. The transcription elongation factor NusA has been identified as capable of terminating transcription in addition to its function as a transcription elongation factor[11]. Another protein, NusG, has been shown to be an important cofactor for Rho-dependent termination and can enhance the rate of RNAP release from the DNA[12,13]. After termination, the core RNAP complex is free to bind a sigma factor and begin the cycle anew.

II. RNAP search process

Once RNAP finds its target promoter sequence, the kinetics of transcription follow a highly prescribed order. However, the search process- or how RNAP locates a promoter sequence to bind to- can be variable in time for individual RNAP. The reaction rate is theoretically dictated by a few factors: the diffusion coefficients of the respective partners (i.e. RNAP and the promoter sequence), and the distance apart the two need to be in order for a reaction to occur. Smoluchowski solved this relationship. Simplifying this scenario by assuming one of the binding partners movement is negligible compared to the other- i.e. DNA moves much more slowly than RNAP- the diffusion limited reaction rate can be defined as:

$$k_{smol} = 4\pi Dba \tag{1.1}$$

where D is the three-dimensional diffusion coefficient of the mobile binding partner (i.e. RNAP), b is the cross-section of the binding reaction, and a is the fraction of the molecular surface of the mobile binding partner that can react with its stationary binding partner. This is considered the upper limit to a diffusion-limited reaction. If we assume that the cross-section of the binding reaction is 0.34nm, or the distance between two base pairs, the reactionable fraction to be ~0.5, and the diffusion coefficient of RNAP to be about $1e^{7}$ nm^{2}/s, this gives an upper limit of ~$10e^{7}$ M^{-1} s^{-1}. For many studied promoters, the observed association rate is usually in this order of magnitude, but occasionally it appears to break the diffusion limit. This garnered the interest of many biophysicists to understand the mechanism of how proteins could exceed the diffusion limit.

III. Search process mechanisms

For a system in which a single target exists, and the molecule of interest has freedom to diffuse in three-dimensions, there is a finite probability that the molecule will never find its target. Infinite waiting times are not conducive for life, therefore nature must have developed a way to avoid such a situation. Adams and Delbruck proposed a mechanism in which the search process is sped up through a reduction in dimensionality. While it is possible to never find a target in three-dimensions, a molecule will always eventually find its target when searching in two dimensions, according to a probability theorem by Polya[14]. However, this theorem doesn't consider the time of diffusion itself; therefore, Adams and Delbruck demonstrate that certain ratios of 3D and 2D diffusion coefficients as well the ratio of the diffusing space to the target space can provide favorable search times[15]. Using a lattice model, Holyst and colleagues

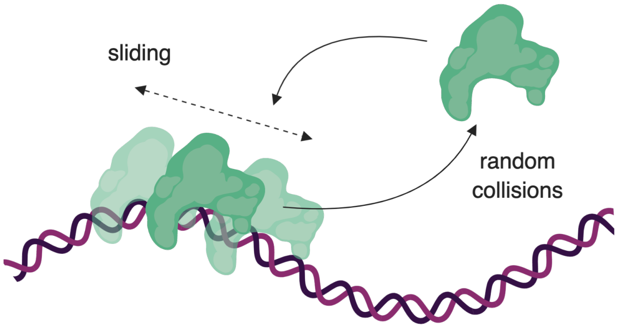


Figure 1.1: model of facilitated diffusion.

A DNA-binding protein can increase its search by combining random collisions with one-dimensional sliding.

independently verify this model can achieve significant increase in search time if the attraction potential to keep the particle diffusing in 2D was high enough[16]. It follows that further reductions in dimensionality would greatly increase the search time.

This lead to the mechanism of facilitated diffusion (Figure 1.1), in which a protein molecule bound to DNA and could slide one-dimensionally for a designated length before unbinding, freely diffusing, and binding DNA again. This mechanism essentially follows a reduction in dimensionality from three dimensions to one dimension, allowing for a faster search time than pure three-dimensional diffusion otherwise would. This model was favored by most, as it was intuitive- DNA can be seen as a linear piece of code, so a protein molecule bound to it would be bound along a one-dimensional tract. Many factors can influence this simple facilitated diffusion model, which I discuss in detail in the following sections.

IV. Factors that influence the search process

A. DNA Roadblocks

As mentioned, DNA within a cellular context is never naked. In prokaryotes, multitudes of nucleoid-associated proteins, transcription factors, repair enzymes, and other DNA binding proteins coat the chromosome. Eukaryotes are even more complex, as their DNA is packaged into nucleosomes and further packaged into chromosome territories and other organization motifs. In addition, eukaryotes have transcription factors and chromatin remodeling enzymes that alter the chromatin structure and accessibility. However, despite the multitude of roadblocks, prior theoretical work by Li and colleagues found that one-dimensional sliding enhanced the search process even in the presence of roadblocks[17]. Confirming this theory, Hammar and colleagues developed a single molecule based single operator binding assay to directly assess the search process of the *lac* repressor in live *E. coli* cells[18]. In this study, the researchers inserted two *lac* repressor operator sites into the chromosome with a variable number of basepairs between the two sites, then measured how fast a fluorescently-tagged LacI protein could find either site. If the sites are such a distance that is equal to or less than the typical sliding length, the search time kinetics should mimic that of a single operator site. Using this assay, they determined that LacI slides ~45bp on average. To probe the effect of DNA obstacles on the LacI search time, they inserted a *tet* operator site with upstream of the *lac* operator, where TetR would bind and create a roadblock, and examined its effects on association kinetics of LacI. They found that the

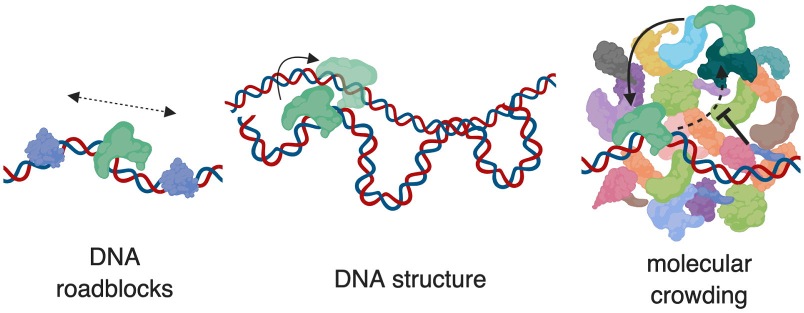


Figure 1.2: a few examples of factors that influence the search process.

A DNA-binding protein undergoing target search on DNA must contend with several factors, such as DNA roadblocks (left), i.e. the multitude of other DNA-binding proteins; DNA structure (middle), the complex three-dimensional structure of DNA due to polymer physics and chromosome organization, as well as the local supercoiling state; and molecular crowding (right), the weak interactions of a DNA-binding protein with the millions of other proteins within the cellular milieu.

association rate increased with the presence of TetR bound near the operator site, confirming the group's prior theoretical work.

Marcovitz and Levy investigated the effects of DNA-bound obstacles on the search process using various computational approaches. Their simulations found that DNA obstacles, even at relatively low concentrations, could hinder the search process by confining particles sliding one-dimensionally on DNA. However, this effect can by circumvented by interspacing these one-dimensional sliding events with hopping events. Increasing the number of obstacles tended towards a respective increase in hopping event frequency. In addition, obstacles on DNA are rarely static, and their results indicate that this dynamic motion of DNA-binding proteins may further enhance the rate of correct target identification[19].

Another theoretical work done by Shvets and colleagues aimed to understand how static and dynamic obstacles of variable sizes and residence times affected the search time. In the case of the static obstacle, the target size, protein sliding length, and length of uncovered DNA were more important than the presence of the obstacle itself. When the sliding length is less than the target size, this essentially corresponds to a pure three-dimensional searching strategy, where the protein binds and unbinds randomly. Thus, association of an obstacle has negligible effect. For the case where the sliding length is greater than the target size but less than the length of uncovered DNA, static obstacles could decrease the search time, but this effect was position-dependent: only when obstacles were directly adjacent to the target site did the search time increase. Interestingly, this result directly contradicts the results from Li and colleagues. Next, the researchers examined the effects of dynamic obstacles. Again, for purely three-dimensional diffusing proteins, addition of an obstacle, even a dynamic one, has no effect on the search time. For molecules whose sliding length is less than the length of uncovered DNA but greater than the target size, the result was similar: because the obstacle is smaller than the total DNA length, and because the protein diffuses relatively quickly, it loses memory of where the obstacle was, thus the only factor that ultimately matters is the length of uncovered DNA. However, modulating the on and off rates of the obstacle could narrow the dynamic range of search times[20].

Ultimately, their stochastic model has some key features that other models such as Hammar et. al.'s continuum model include, but they argue their model is more realistic by examining the extreme cases- i.e. when the sliding length becomes infinite, the continuum model suggests the search time will be twice as long[18], but their model suggests the search time should approach infinity. Their reasoning is that when a molecule has increasing sliding length, it can get trapped in the blocked segment of the DNA away from the target and, if the sliding length is infinite, will never dissociate and reach the unblocked DNA segment, increasing the average search time to infinity. They note that while their model is more physiologically relevant, it does not take into consideration a multitude of other physiologically relevant factors such as the movement of the obstacle on the DNA (thus occasionally obstructing the target), using non-linear DNA, effects of proteins interacting with crowding reagents, and sequence-dependent effects[20].

B. DNA structure

In addition to roadblocks, DNA structure can have an effect on the search process. DNA does not exist in cells as a linear line, but compacts into spatially segregated macrodomains that have complex, three-dimensional structures. Using an optical tweezers approach, Broek and colleagues measured the association rate of the restriction enzyme EcoRV to DNA as a function of DNA extension. By extending the theoretical framework of Berg and Ehrenberg, the authors quantified the effect of "intersegmental jumps", i.e. the transfer of the DNA-binding protein from one site of DNA to a physically close but genetically distant site, on the search process. Their data indicated that DNA coiling was beneficial to the search process. In their argument, they proposed two extremes.

The first is a case where the three-dimensional diffusion coefficient is much larger than the association rate. In such a case, the protein undergoes large excursions in three-dimensional space before binding to DNA, an inefficient search strategy. The other extreme is the case where the three-dimensional diffusion coefficient is much smaller than the association rate. Here, the protein would dissociate from the DNA and have a high probability of binding to the same site, grossly oversampling the search space. With DNA coiling, the dynamics of DNA are such that when a protein unbinds from the DNA, a random DNA segment is located near its previously bound site, increasing the probability to bind the random DNA segment over the previously bound site and substantially reducing the oversampling rate [21]. Soon after this work was published, Lomholt and colleagues, from the same lab, developed a more sophisticated theoretical model to describe the effects of DNA coiling on the search process. Using an analytical framework that does not rely on treating DNA as cylinders, they again demonstrate the acceleration effect of inter-segmental jumps on the search process [22].

Zhou and Szabo developed an analytical framework to monitor the search process of proteins to DNA sites by allowing proteins to bind non-specifically to DNA. The entire binding surface for a diffuser was treated as containing a short-range attractive potential, which they postulate is more physiologically relevant than one-dimensional sliding. Using this framework, they demonstrate that such non-specific binding to an attractive surface increases the association rate to a specific site. Interestingly, their framework also implies non-exponential relaxation of bound molecules[23]. Other studies have proposed this type of behavior due to various physical effects, which I will discuss in a section below.

In another theoretical work, Mahmutovic and colleagues probed which effects- sliding, hopping, intersegmental transfer- was most important for the LacI search process. They found that sliding was the most crucial factor, and hopping, even from one site to a nearby random DNA site, had little effect on the search time, as it hindered one-dimensional sliding[24]. However, they note that proteins that do not use facilitated diffusion would benefit greatly from hopping, as proteins that were close to the target site would bind more quickly rather than having to restart the entire search process. This hints at the possibility that different proteins may have different search strategies. Indeed, a study examining two restriction enzymes- EcoRI and EcoRV- found that each displayed different search strategies, with EcoRI favoring one-dimensional sliding while EcoRV favored hopping and intersegmental jumping[25].

In 2009, work done in the Ellenberg lab demonstrated direct *in vivo* evidence for the effect of DNA structure on the search process of nuclear proteins. Using photoactivation and monitoring the fluorescence out of the photoactivated site (principally similar to the FRAP technique), they found binding of nuclear proteins such as the H1.1 histone subunit to DNA was enhanced within heterochromatin. Using FCS to monitor various fluorescent proteins, and tracking individual injected quantum dots, they determined the enhancement of protein association within heterochromatin was due to the fractal nature of the chromatin itself[26] (Chapter 1, Section VII.A). Essentially, the structure of the chromatin itself acted to limit the space molecules could search, resulting in what the literature calls “compact exploration”. Molecules that diffuse within the heterochromatin have a greater chance of staying within the heterochromatin structure; this effective trapping of the molecule accelerates the search process, as the molecule has much fewer sites to visit. This work gave credence to the idea that the fractal structure of the chromosome could influence protein diffusion and gave direct *in vivo* evidence of the subsequent consequences of this type of diffusion on searching kinetics.

C. Molecular crowding

Crowding is an often neglected, but critical component for nearly all interactions within the cell. The importance of molecular crowding in biochemical reactions has been known since 1963, wherein Laurent performed a systematic study of the effects of dextran on protein solubility; in constant high ionic strength media, increased concentrations of dextran lead to a decrease in protein solubility for serum albumin[27]. In addition to protein folding, protein-protein interactions, including oligomerization, were found to be enhanced under crowding conditions. One of the first studies to observe this phenomenon found that replication of an *oriC* plasmid in a cell-free system could only be achieved by adding the crowding reagent polyethelyne glycol (PEG)[28]. The corresponding author, Dr. Kornberg, even added that molecular crowding must be corrected for when working in dilute conditions as part of his scientific “ten commandments”. Many studies afterwards had similar observations. For example, α-synuclein was found to form aggregates over a very extended period in dilute conditions, but in a much shorter time frame under crowding conditions[29]. Batra and colleagues, who studied the effects of crowding on the association of the σ and ε subunits of *E. coli* DNA polymerase III, more carefully characterized the thermodynamic effect of crowding on protein-protein interactions. Their studies found that the stabilizing effect is relatively modest in terms of thermal energy- around 1 kcal/mol- but that these stabilizing effects are cumulative, thus leading to substantial stabilization of large protein complexes[30].

These and many other studies highlighted the substantial effects of molecular crowding on the kinetics of nearly every biochemical reaction inside the cell. This of course does not exclude the search process of proteins to their DNA targets. In general, crowding has three major effects: (1) hindrance of diffusion[31], (2) anomalous diffusion[32,33], and (3) increased protein association[34]. An early *in vitro* study measured the effect of crowding agents on the kinetics and binding of EcoRV by adding the crowding reagent Ficoll and measuring the properties of the cleavage reaction- i.e. K_m, V_{max}, and the non-specific dissociation constant. Their results found that crowding did influence the dynamics by increasing both K_m and V_{max}, but these increases were offset by a decrease in the non-specific dissociation rate. Thus, while crowding has a nonzero effect, the net result of crowding is negligible in terms of binding and cleavage[35].

A later theoretical work attempted to formalize the effects of crowding environments on Michaelis-Menton enzymatic reactions through Monte Carlo simulations. They use a lattice model where the enzyme, substrate, complex, and product can move on the lattice but are hindered by obstacles placed such that the free space for movement describes a specific type of fractal called a percolation cluster (see Chapter 1, Section VII.A), chosen to mimic the crowding of the cytoplasm or cell membrane. Strikingly, their results show that this simple model recapitulates so-called “fractal kinetics”, in which the coefficient which relates the rate of the reaction to the concentration of the reaction components is no longer constant, but instead follows a power law through time. They show the effects of different obstacle densities on the rate; overall, increasing obstacles had little effect on the rate over a short time interval but substantially decreased the reaction rate over time. Most interestingly, their lattice showing the locations of the enzyme and substrate demonstrated that at higher obstacle densities, both enzyme and substrate self-separated over time, which may explain the reduction in reaction rate at later time points. Importantly, their work shows that for molecules diffusing within a fractal environment, steady state assumptions may not be valid[36].

Following up on this study, Grima and Schnell used two types of lattice models and an off-lattice model to monitor the effects of molecular crowding on the simple biochemical reaction A + B → C. Using Monte Carlo simulations of diffusion, they monitor the reaction rate of the formation of C over time for a variety of obstacle concentrations in the range predicted by biological studies (i.e. 1 to 50%). They argue that previous theoretical studies demonstrating fractal kinetics found the effects to only occur at long timescales and may not be relevant in biological contexts. Thus, they only monitor the reaction until it is 99% completed. Interestingly, they find that both classical mass action and fractal kinetics can occur. In the case where the reaction probability is low, or the initial concentration of A is much lower than the initial concentration of B, the kinetics closely match classical mass action, and thus it can be safe to ignore the effects of volume exclusion. However, they note that in the case of a heterogeneously mixed population with a high concentration of obstacles, the reaction is best described by fractal-like kinetics. This study underscores how multiple factors: reaction probability, initial concentrations, sizes of the particles, mobility of the particles, and the spatial geometry of the surroundings- can have an influence on the extent that the kinetics deviate from classical mass action[37].

Perhaps unsurprisingly, crowding has a dramatic effect on the search process of proteins. Cravens and colleagues found that crowding greatly impacted the search of two human glycosylase enzymes *in vitro* by increasing the contact time of the enzyme with DNA, and increasing the length it slides/hops one-dimensionally, demonstrating that the physical properties of crowding has as large an impact on the search process as the physical properties of the protein[38].

In addition to its effects on kinetics, crowding can dramatically affect protein diffusion. In theory, the chemical potential only relies on the temperature and concentration of the molecules in the solution. However, in crowded environments, these molecules have weak interactions with their neighboring molecules, leading to a non-linear increase in chemical potential based on the size and number of molecules contributing to crowding[39]. As such, this leads to long-range distance correlations, described by the fractional Brownian motion model. In addition, the cytoskeleton or DNA of the cell can act as static obstacles, as mimicked in the lattice models above. Moreover, anomalous diffusion could arise from dynamic crowders that "trap" molecules within localized cages with waiting times that follow a power law; this type of motion is called continuous time random walk (CTRW). None of these models are mutually exclusive, but each leads to different behaviors and implies different biological mechanisms responsible for anomalous diffusion.

Elowitz and colleagues first demonstrated the effect of crowding on protein diffusion through FRAP experiments. Their results found the diffusion coefficient of green fluorescent protein (GFP) to be an order of magnitude less than predicted by the theory at the time. Additionally, addition of a 72 kDa protein to GFP, which hypothetically should have had only a marginal effect on the diffusion coefficient, impeded the diffusion more than two-fold. Most strikingly, excessive overexpression of GFP also decreased the diffusion of GFP by two fold[31]. While seemingly puzzling, these results are consistent with the idea that molecular crowding contributes to diffusion in a non-trivial manner dependent on the respective sizes of the diffuser and adjacent macromolecular components and their interactions.

Weiss and colleagues used fluorescence correlation spectroscopy experiments coupled with simulations to demonstrate that molecular crowding indeed leads to anomalous diffusion (i.e. mean squared displacements that follow a power law through time with exponent less than one). Additionally, they demonstrate the anomalous diffusion exponent gives a quantifiable measure for the extent of molecular crowding within living cells[40]. Golding and Cox later performed single particle tracking experiments (see Chapter 1, Section VII.B) on single RNA molecules in *E. coli* cells, demonstrating that individual molecules display anomalous diffusion, again likely due to molecular crowding within the *E. coli* cytoplasm[32]. Clearly, crowding greatly influences the diffusion and kinetics of molecules and undeniably affects the search process of DNA-binding proteins to their DNA targets.

V. Studies of the search mechanisms of DNA-binding proteins in live cells

Renewed interest in the effects of crowding on the search process has recently surged in the field, thanks to the development of superresolution microscopy. Initially developed to examine the structure of protein complexes such as actin networks within cells, it was modulated to be able to do high-resolution single particle tracking (SPT) of individual proteins within live cells. SPT is a powerful technique that is able to monitor the dynamics of individual proteins in time (see Chapter 1, Section VII.B for technical details), gaining useful information such as diffusion coefficients of molecules and their diffusive state switching kinetics, in addition to how and where molecules diffuse.

In seminal work, Elf and colleagues used fluorescence microscopy to monitor the dynamics of the *lac* repressor in real time. To measure the dissociation rate of the *lac* repressor from the *lac* operator site, they added varying concentrations of IPTG, a mimic of allolactose that induces an allosteric conformational change to LacI that prevents it from binding the *lac* operator site, allowing for transcription from the *lac* promoter. To measure the association rate after all LacI molecules had dissociated under the high concentration of IPTG, they diluted the IPTG concentration by adding media with a high concentration of ONPF, a competitive anti-inducer that prevents IPTG from rebinding once dissociated. Finally, they use stroboscopic SPT to monitor non-specifically bound LacI molecules and determine their diffusion coefficient to be ~1 order of magnitude greater than specifically-bound LacI molecules, which they theorize to be LacI binding interspersed with random 3D diffusion. Interestingly, they find the diffusion to follow normal Brownian motion and not subdiffusion, as had been previously found with mRNA molecules[32]. Using FCS, they estimate the 3D diffusion coefficient to be ~1 order of magnitude greater than the non-specific binding diffusion coefficient leading to an estimation that ~87% of LacI molecules are bound non-specifically to DNA at any given time. Combining this information together with a simple facilitated diffusion model, they were able to estimate the search time to be less than 270s for a single LacI molecule to find a single *lac* operator site[41].

Izeddin and colleagues used SPT to examine the search strategies of two proteins, the transcription factor c-Myc and the transcription elongation factor P-TEFb. To examine their diffusive behavior in more detail, the researchers measured the angle between two consecutive displacements for millions of trajectories and plotted the histogram of these angles. For purely Brownian motion, this histogram should be even; i.e. the probability for a molecule to diffuse in any given direction should be identical for all angles. However, if the molecule experiences confinement or crowding, this prevents forward-step angles from occurring, and thus the histogram reveals a bias towards backward-step angles. Interestingly, they found that c-Myc displayed normal diffusion, while P-TEFb displayed anomalous diffusion. Likewise, c-Myc showed mostly homogenous distribution of displacement angles, while P-TEFb demonstrated a heavy bias towards backwards angles. Using simulations, they were able to demonstrate that the behavior of P-TEFb is consistent with a molecule diffusing and becoming transiently trapped for a variety of waiting times that follow a power law (i.e. a CTRW model), implying the spatial geometry of the nucleus is important for determining output diffusive behaviors. They conclude that c-Myc is a “global explorer” of the nucleus, while P-TEFb is a “local explorer” that oversamples its environment. Consequently, c-Myc is able to bind its target anywhere in the nucleus, while P-TEFb locates its target in a position-dependent manner. They conclude that the interactions of the two transcription factors with the nuclear environment, not their exclusion from certain compartments, dictate their search strategies and kinetics[42].

Normanno and colleagues investigated the effects of non-specific binding on the search process by monitoring dye-labelled TetR injected into the nucleus of U2OS cells with a fluorescent-reporter operator system upstream of a *tet* operator site inserted into the genome. By labeling the correct target site, the researchers were able to monitor TetR binding to non-cognate sites in its target search. Using a rolling-window average, they were able to monitor the diffusion in time and determine the DNA-bound and freely diffusing states, and consequently measure the residence times on non-specific DNA. Strikingly, they found that the distribution of residence times followed a power law instead of the predicted exponential decay from normal biochemical rate laws. This would suggest non-specific binding as a rate-limiting step in the search process. To circumvent this, TetR spends much of its time freely diffusing in solution with only transient associations with DNA[43]. The behavior of LacI and other native eukaryotic transcription factors were found to be similar in behavior. This is interesting given that in *E. coli*, as demonstrated by Elf and colleagues, LacI spends most of its time bound to non-cognate DNA[41]. This implies that the physical properties of the cytoplasm/nucleus itself may greatly influence the search strategy of DNA-binding proteins. This idea is consistent with the study by Izeddin and colleagues[42].

Normanno and colleagues found the distribution of residence times of TetR on non-specific DNA followed a power law. The authors contributed this to TetR binding to near-cognate DNA, increasing the time bound due to near matches in cognate binding. However, it can be argued that this would still lead to an exponential distribution in non-specific binding times; if each unique sequence that TetR binds produces a unique exponential rate, the sum of these rates in a biochemical rate reaction would give an effective rate equal to the weighted sum of all of these off rates with the respective concentration of their unique sequences. Zaid et. al. demonstrate using a continuous time random walk model that unbinding from a surface follows a power law related to the power law of the diffusive behavior[44]. Interestingly, the model by Zhou and colleagues, in which they applied a simple attractive field to the modeled DNA surface, also found non-exponential distribution of non-specific binding times[23], suggesting that non-uniform weak interactions, no matter the underlying physical cause, leads to dramatic effects on search kinetics. This result is in line with prior theoretical studies demonstrating the breakdown of steady state assumptions under the effects of crowding[36].

VI. Search mechanism of RNAP

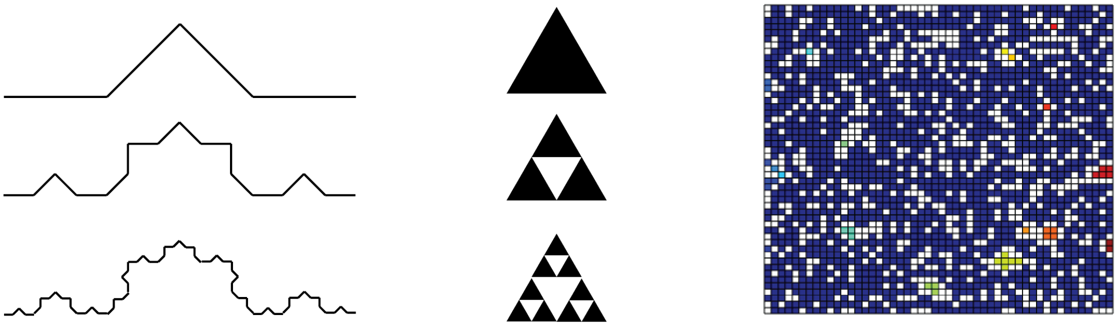
Recent studies of the search process of *E. coli* RNAP have shown that it does not find its target through facilitated diffusion, but rather by random collisions with DNA. Friedman and colleagues demonstrated this by adhering fluorescently labeled DNA molecules containing a single promoter site and a variable number of base pairs upstream of the promoter site. If RNAP underwent one-dimensional sliding, adding base pairs upstream of the promoter should enhance the RNAP-promoter association rate, determined by the rate of colocalization of fluorescently labeled RNAP with the fluorescently labeled DNA. Instead, Friedman and colleagues found that the association rate did not change with increasing number of upstream nucleotides, indicating that RNAP finds its target purely by random chance from three dimensional diffusion, with no one-dimensional sliding[45].

Wang and colleagues confirmed this observation in a parallel study. Using fluorescently labeled DNA curtains, fluorescently labeled RNAP, and fast tracking, they were able to directly assess the search process of RNAP. They found that under a certain concentration of RNAP well below the physiological concentration, RNAP could undergo facilitated diffusion, but otherwise found its target promoter through pure three-dimensional diffusion. Interestingly, their work also monitored the search process of LacI to its operator site and found it to generally follow the same principles, though with different concentration requirements for the switch to pure three-dimensional diffusion. Wang and colleagues developed a theoretical framework in which to explain these concentration-dependent results. Briefly, they postulate that facilitated diffusion and three-dimensional diffusion binding were separate and competing mechanisms. When the concentration of the protein is increased above a certain point, three-dimensional diffusional binding dominates due to the increased probability of the protein randomly colliding with the target sequence[46]. This may explain the relative differences in search strategies between LacI and RNAP, as LacI is generally expressed in low quantities (~20nM, likely below the threshold to switch to 3D-diffusion dominated searching) while RNAP is highly expressed (~10uM, well above the threshold).

The authors note that their imaging conditions are in dilute settings, and that factors such as other DNA-binding proteins, macromolecular crowding, and transcription factors could influence the kinetics of the search process. For instance, there exist a class of proteins called nucleoid-associated proteins (NAPs) that are responsible for nucleoid organization. Many of these proteins, such as HU or Fis, are expressed in extremely high abundance (~10x higher than RNAP, or 100-200 uM) and may coat DNA and affect the search process, by eliminating non-cognate sites or by obstructing the target site. Additionally, as previously demonstrated, crowding has a substantial effect on protein diffusion and reaction kinetics, and thus will inevitably alter the parameters determined in this *in vitro* study. Such effects highlight the need for studies detailing the search process of RNAP within a biological context. Because of the importance of gene regulation at the level of transcription in *E. coli*, it is pivotal to understand the search process of RNAP within the cytoplasm to gain a realistic picture of gene regulation dynamics.

VII. **Supplementary Information**

A. Fractals



Koch curve

iteratively insert a triangle
into every straight line segment

Sierpinski gasket

iteratively subtract a triangle
from the center of each black triangle

percolation cluster

two-dimensional lattice where each
site is occupied with probability p
Dark blue is the largest cluster which
extends end to end

Figure 1.3: example of different types of fractals.

Both the Koch curve and the Sierpinski gasket are examples of fractals that are recursively generated and thus are exactly self-similar at different length scales. A percolation cluster is a randomly generated fractal that is statistically self-similar at different length scales across the entire lattice. All fractals show scale-invariance in structure and thus scale invariance for diffusion of particles within them.

Fractals are any type of object that demonstrates self-similarity over a range of distance scales. Generally, they are irregular and thus difficult to define using classical Euclidean geometry, is at least statistically self-similar at different length scales, and can be defined using a simple recursive function. Some example fractals and their descriptions are given below in Figure 1.3. Fractals gained the interest of biophysicists as diffusion within a fractal environment has interesting effects. In non-fractal, non-crowded environments, molecules undergo traditional Brownian motion; i.e. random movement correlated with the environmental temperature and the molecule’s size. The average displacement of any given molecule over time will be zero due to the total randomness of the motion. Plotting the mean squared displacement over time, however, gives a linear line in which the slope of the line directly correlates to the diffusion coefficient of that molecule. In the case of a molecule diffusing within a fractal environment, the motion becomes non-linear, instead following a power law with exponential less than 1 that is directly related to the fractal dimension ($\alpha = (1+v)/2$ where α is the power exponent for the anomalous diffusion and v is the fractal dimension along one dimension). This relationship demonstrates that with increasing fractal complexity comes a decrease in the power law exponent of the mean squared displacement of diffusing molecules.

B. Single particle tracking

Single particle tracking involves the tagging of proteins or other macromolecules with a fluorescent molecule- usually either a fluorescent dye or a fluorescent protein- and using a fast camera speed to monitor the dynamics of proteins in real time. Like STORM or PALM, the spatial positions of molecules in each frame are determined by fitting pixels of high intensity to a two-dimensional Gaussian, which approximates the Airy disk point spread function generated by a point source of light. Such fitting reduces the uncertainty in a molecule's location from 200nm or so to ~20-40nm. Individual detections are linked through space and time by thresholds; i.e., molecules close in space and time are generally linked together, while those that are far apart in space and separated by a long time interval are generally not. Many algorithms have been developed to better

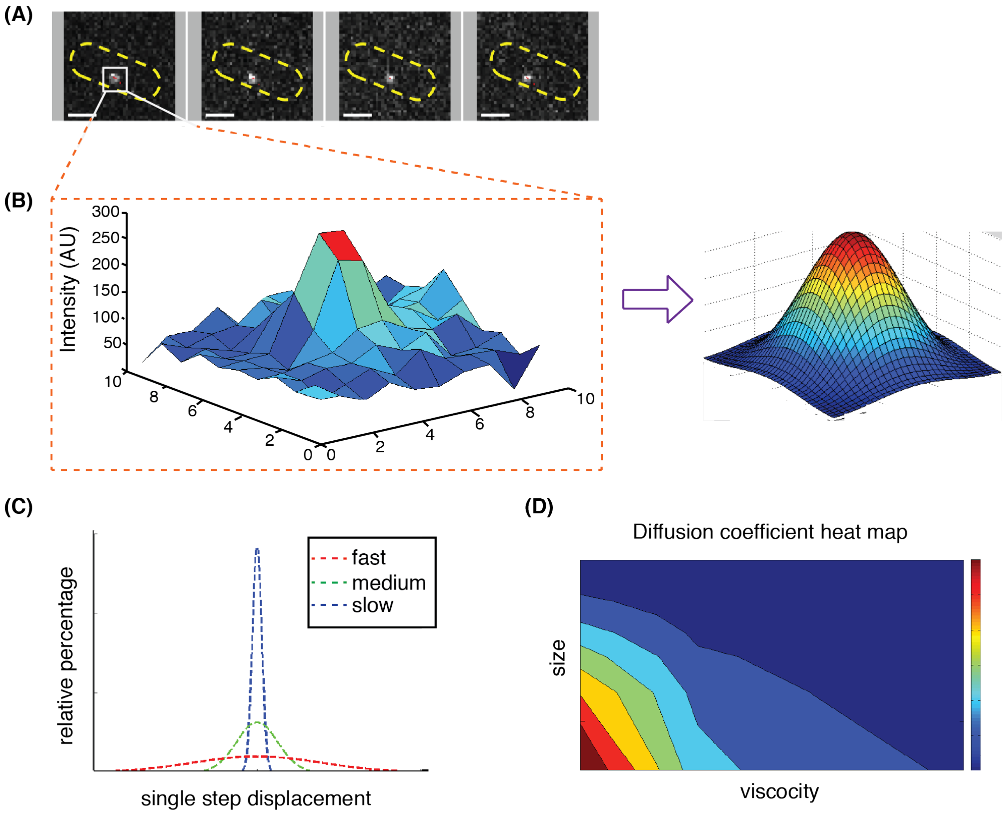


Figure 1.4: brief explanation of single particle tracking.

(A) Montage of single molecules within bacterial cells collected in time. (B) Individual molecule detections are isolated and fit to a two dimensional Gaussian. (C) The "speed" of diffusion changes with diffusion coefficient; fast = larger diffusion coefficients and slow = slower diffusion coefficients. (D) Heat map of the change in diffusion coefficient as a function of fluid viscosity and particle size. Blue indicates slower diffusion coefficients while red indicates faster diffusion coefficients.

link molecules together, generally by adding in a cost matrix or through a maximum likelihood approach[47]. The most important parameter calculated from SPT is the diffusion coefficient, which is defined as the ratio of the flux of the molecule in a certain cross-section over the gradient in the concentration of the molecule. The driving force for diffusion is through its chemical potential-molecules tend to spread themselves homogenously in solution. Colloquially, the diffusion coefficient describes the "speed" at which a molecule moves (diffusion is not technically a speed as it is expressed in units of length squared over time), or how quickly a molecule can span a given area. The diffusion coefficient of any particular molecule is dependent on the molecule’s size (with the longest diameter having the greatest effect), the viscosity of the surrounding media, and the temperature of the environment. Determining a molecule’s diffusion coefficient can therefore give information about its local environment. For example, molecules that bind to other molecules experience an effective increase in size, and therefore a slower diffusion coefficient. By monitoring how often molecules switch between a fast diffusion coefficient and a slow diffusion coefficient, one can determine the *in vivo* binding kinetics of a given protein. There are many methods of calculating the diffusion coefficient(s) of a molecule from SPT data, which are reviewed in Vrljic, et. al.[48], and newer methods based on hidden Markov models[49,50].

VIII. Concluding Remarks

With more sophisticated single molecule tracking technologies and analysis algorithms, the effects of molecular crowding on protein dynamics is becoming increasing more appreciated. With the advent of membrane-permeable dyes, it is now possible to track individual molecules on the minutes timescale, orders of magnitude better than nearly all modern fluorescent proteins. This will provide rich information on molecular movement, allowing experimenters to directly assess correlations of movement over timescales over multiple magnitudes. This has been done for DNA loci, for which it has been previously possible to monitor over long periods of time. The motion of DNA is indeed not simplistic and shows different types of subdiffusive behavior over differing time regimes.

It is likely that this anomalous diffusion is not the result of only one of the proposed mechanisms (fractal obstacles, CTRW, or fBM), but rather is a result of all three. Several models of the chromosome, from bacteria to humans, have proposed a fractal structure of some kind. As DNA is a relatively immobile molecule compared to nearly all other proteins in the cell, this would manifest as a static obstacle. In addition, the density of proteins cannot be understated, and protein molecules must be mobile to facilitate life; thus, it is not a stretch to imagine that molecules may cage other molecules for a wide distribution of times akin to the CTRW model. Lastly, as many molecules have weak binding affinities for the plethora of other molecules present in the cell, one could imagine that this creates a complex intracellular environment with non-trivial forces acting upon individual molecule. Simplistically, one could imagine that the chemical potential of the molecule is not only a function of the concentration of the molecule itself, but the sum of the potentials from all other molecules in within the cellular environment. This might manifest as long-range “forces” that create a sort of “memory” on the molecule leading to long-range correlations in displacements seen in fractional Brownian motion. Determining which of these models are relevant *in vivo* will not be trivial. Schulz and colleagues derive a correlated CTRW model that encompasses the main features of these three types of possible mechanisms and demonstrate that large variations in parameter space lead to probability density functions that are indistinguishable from each other[51].

Regardless of the physical mechanisms underlying the motion, several recent studies have demonstrated the non-classical kinetics of protein search in live cells. Understanding the extent and/or timescales these deviations from classical mass action occur will have important implications for gene expression. Additionally, such parameters would be extremely useful for synthetic biology as a reference for standard cellular behavior. It is clear that proteins have evolved for different search mechanisms, and so far, the physical properties of individual proteins in relation to their environment that facilitate their differing search mechanisms remains largely unknown. It is clear that the search strategies are not inherent to the protein structure itself, as TetR and LacI show distinct search patterns in eukaryotic cells over the bacterial cells from which they originate. Teasing out such relations will be non-trivial but would greatly improve our understanding of the diversity of protein structure and function.

In summary, it is clear that the cellular environment has profound effects on the biochemistry of the cell, with effects arising mostly from macromolecular crowding. As imaging technologies become increasingly more advanced, I predict that an advent of new studies will arise in which biochemical studies are not done *in vitro* but rather are done within living cells.

Chapter 2: Characterization of the kinetics of transcription in live *E. coli* cells

I. Introduction

Gene expression and gene regulation of *E. coli* in response to environmental stimuli relies on the ability of RNA polymerase to navigate the crowded environment of the cytoplasm to find its target in a responsive, efficient manner. Bacteria have a single RNA polymerase (RNAP) that is responsible for all transcription in the cell and utilize transcription factors called sigma factors to enable RNAP to transcribe a subset of genes (reviewed in [1]). Core RNAP, which comprises the subunits $\alpha_2\beta\beta'\omega$ has been shown in biochemical assays to have poor affinity for DNA and is unable to recognize the specific promoter DNA sequences. The holoenzyme, or $\alpha_2\beta\beta'\omega\sigma$, on the other hand, has very high specificity for specific promoter sequences conferred by the sigma factor itself[3]. Thus, to initiate transcription, core RNAP must bind to a σ factor to form the holoenzyme that can undergo three-dimensional diffusion and associate with its σ factor-conferred promoter DNA sequences. Given the high copy number of sigma factors in the cell[5], it has been assumed that RNAP nearly always exists as the holoenzyme, except during transcription elongation.

A typical RNAP first undergoes promoter search to bind a promoter sequence. After binding, RNAP unwinds DNA downstream of the promoter, and begins to synthesize RNA[52-53]. Because of the strong association with the promoter, the rate-limiting step to begin transcription is promoter escape[54]. Until the RNAP is able to dissociate from the promoter, it undergoes a process called abortive initiation, in which 8-10 nucleotides are synthesized, then removed and degraded, wherein RNA synthesis can begin again[6]. After promoter escape, there is a stochastic probability of the sigma factor dissociating from RNAP[8], and several elongation factors such as NusA associate with RNAP to ensure processive transcription elongation[9]. Termination of transcription occurs through one of two mechanisms: Rho-dependent or Rho-independent termination[10]. Once unbound, RNAP is free to bind another σ factor and begin the process anew. While extensively studied via biochemical assays, these studies ignore the crowded, complex environment of the cell. *E. coli* are particularly dense; it has been shown that GFP diffuses two to three times slower in the cytoplasm of *E. coli* as in eukaryotic systems[31]. Additionally, it has been estimated that the total protein volume fraction is ~ 25%[55]. Crowding can substantially influence dynamics of a number of biophysical and biochemical processes, such as protein folding[34], protein-protein interactions[29], and the search strategies of DNA-binding proteins[26,42]. Additionally, it has been determined that mRNAs display subdiffusive behavior in live *E. coli* cells[32]. Thus, it is not a leap to assume the density of proteins within the *E. coli* cytoplasm may affect the transcription cycle kinetics of RNAP.

We utilize a fluorescent-protein tagged RNAP and use fluorescence recovery after photobleaching (FRAP) to study the dynamics of the transcription cycle within the context of the crowded *E. coli* cytoplasm. Using a modeling approach, we determine the recovery can be explained by a simple three state model based on the initiation-elongation-transcription cycle where the best fit parameters match closely to values reported in the literature. To probe the search mechanism of RNAP, we use high-resolution single molecule tracking in combination with statistical modeling to provide a detailed dynamic model of the RNAP search process in live cells. We utilize genetic manipulations and drug treatments to perturb the transcriptional state of the cell in predictable ways to verify the diffusive mode of each state, as well as the kinetics of diffusive state switching. We find that RNAP has defined diffusive modes, corresponding to DNA-bound, rapid association and dissociation within the nucleoid, and freely diffusing. We observe that the diffusive behavior of each state is affected by molecular crowding by confining its movement, and that the non-specific binding time is likely the rate-limiting step in the promoter search process. Lastly, we find evidence that the non-specific binding times are not exponentially distributed, likely again a consequence of molecular crowding. We propose a model in which the limiting steps of non-specific binding are overcome by local confinement of RNAP, limiting its search area.

II. Results

A. FRAP shows complex, slow recovery

To probe the kinetics of the transcription cycle, we employed the well-known methodology FRAP using a fluorescent protein GFPuv labeled RNAP in live MG1655 *E. coli* cells growing in rich defined medium (EZRDM) at room temperature (RT). In this strain, the gene encoding GFPuv is fused to the C-terminus of the rpoC gene encoding the β' subunit of RNAP, and inserted at the native locus on the chromosome to replace the endogenous rpoC gene. We

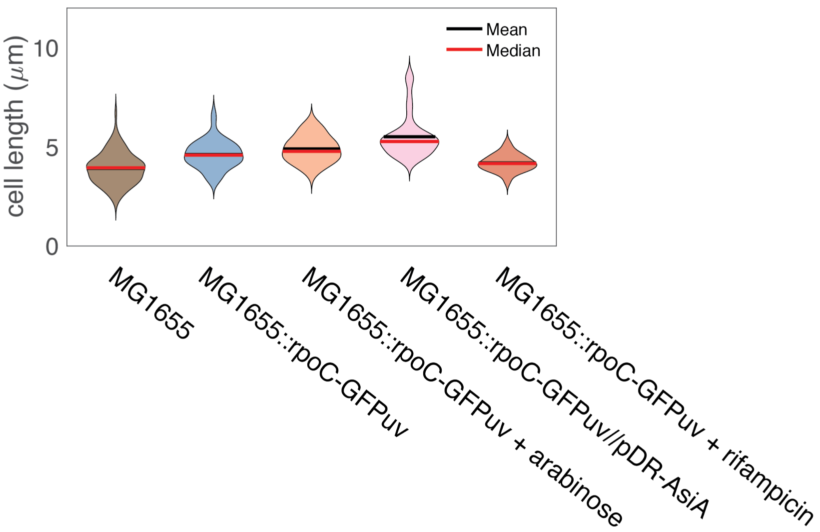


Figure 2.1: comparison of cell lengths of the WT parental strain with the rpoC-GFPuv fusion strain.

Cell lengths from ~ 45 cells in each group are plotted where the width of the bar indicates the relative fraction at a particular cell length value. Black bars represent the mean while red bars represent the median cell length.

show that the rpoC-GFPuv fusion protein had similar cell lengths as WT (Figure 2.1). Thus, the rpoC-GFPuv fusion protein functions fully as the sole cellular source of the β' subunit. In the text below we use the term RNAP-GFPuv to refer to our labeled RNAP.

To perform the FRAP experiments, we imaged individual cells under low 488nm light. Because of the heterogeneity in the distribution of RNAP (Figure 2.2, [56]), we imaged cells for 20 frames to establish the steady state distribution of RNAP. We then used a confocal beam at 100 times the power of our epifluorescence beam to bleach a portion of the cell (a circle ~600nm in diameter) and monitored the recovery of the fluorescence in this bleached region every 1s for a total of 600s (10 minutes). As a control, we also examined RNAP-

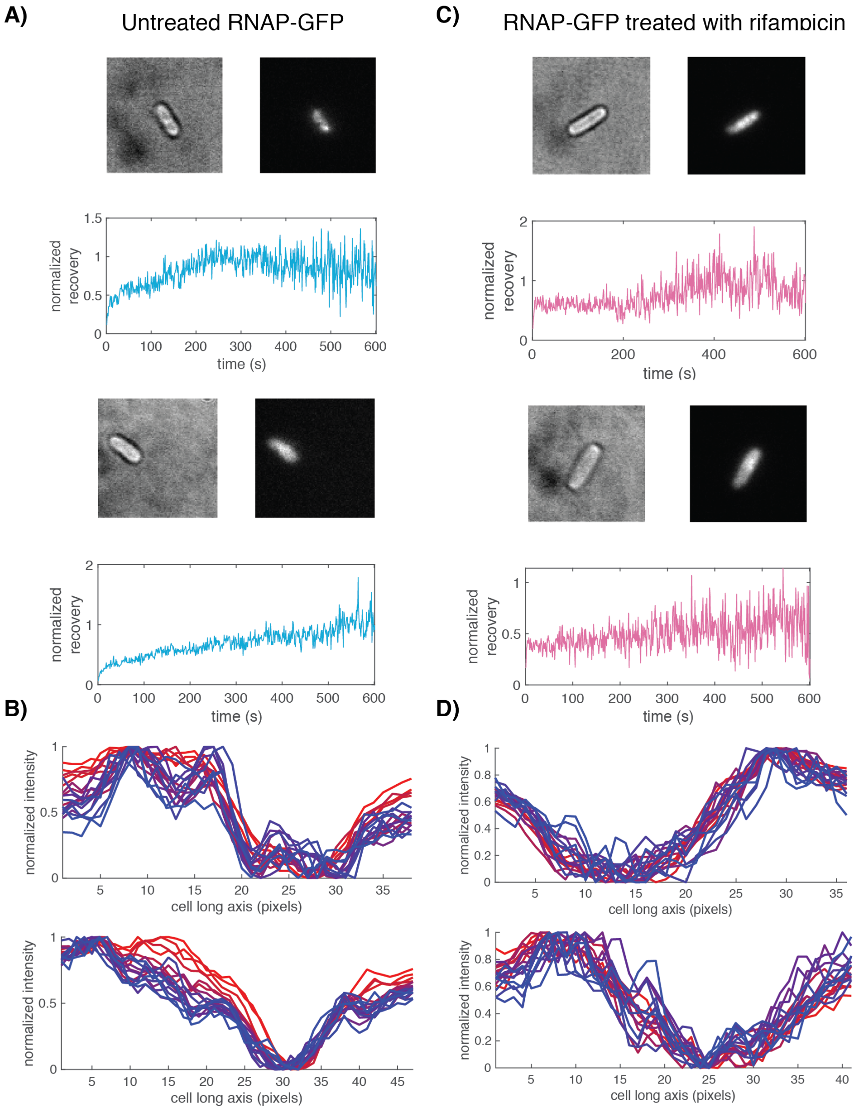


Figure 2.2: FRAP results demonstrate heterogeneous fluorescence recovery that does not depend on diffusion.

(A) Examples of individual cells before photobleaching (i.e. their steady state distributions) and their respective normalized recovery curves. (B) Fluorescence profile of the cell during fluorescence recovery maintains a similar shape over time, indicative that diffusion does not contribute substantially to the recovery dynamics.

GFPuv of cells treated with the transcription inhibitor rifampicin, which binds the RNA exit channel of RNAP, preventing RNAP to undergo promoter escape[57]. As a result, RNAP are stuck on promoter sequences, unable to escape the abortive

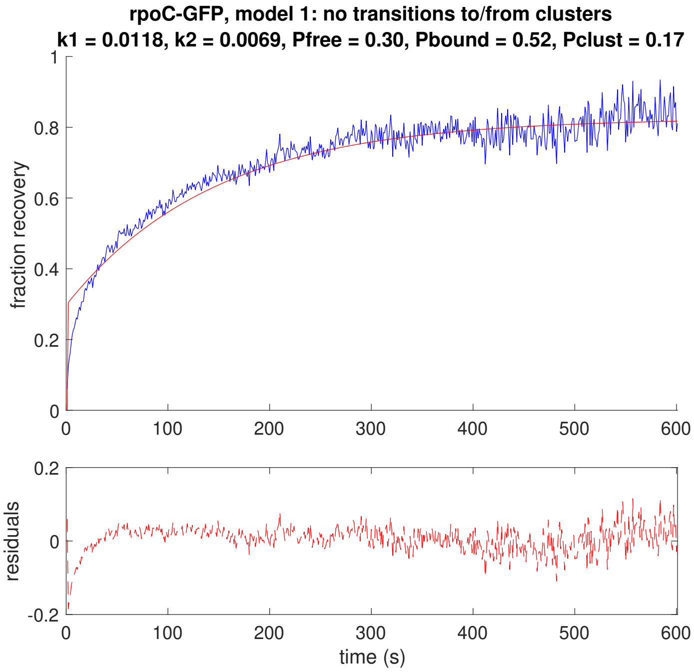


Figure 2.3: FRAP data fit to a simple two-state model does not describe the recovery well.

The average normalized recovery is plotted as a function of time (blue). The red line depicts the best fit model for the simple two state model. Residuals are plotted below, demonstrating the poor fit especially at early timepoints.

initiation phase. Consequently, we should see slow recovery of RNAP in this condition. For each condition, we averaged all of the individual cell recovery curves to prevent misinterpretation due to heterogeneity in RNAP distribution and subsequently heterogeneity in the FRAP recovery curves. Additionally, we show that we can ignore the effects of diffusion on the recovery (Figure 2.2, [56]), validating our model assumptions.

In the WT condition, a two-state model poorly described the recovery (Figure 2.3). We reasoned that RNAP has a designated kinetic cycle, first undergoing promoter search by binding and unbinding non-specific DNA before

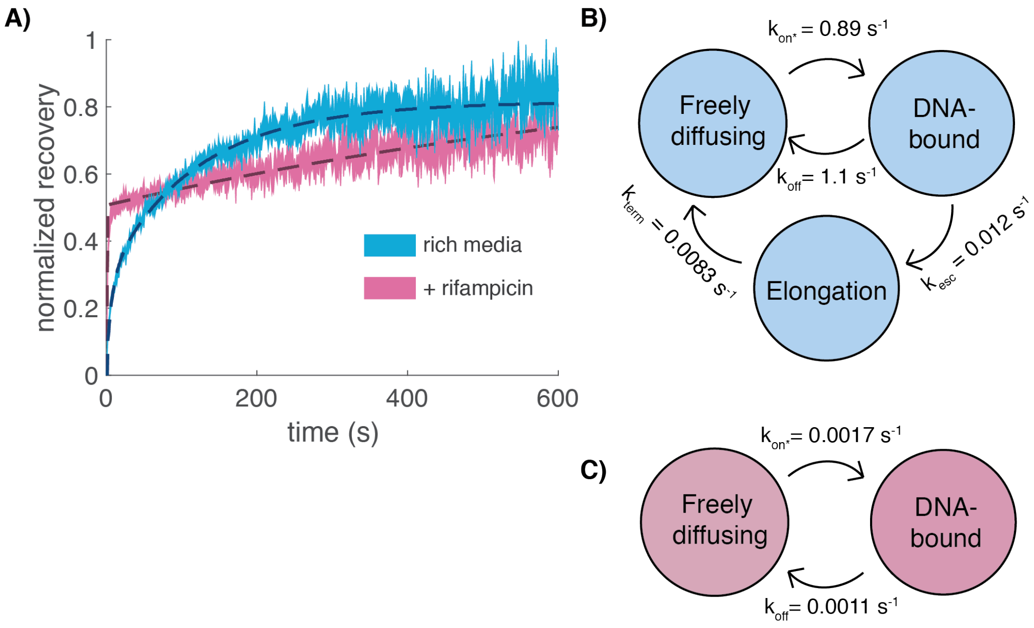


Figure 2.4: FRAP data in EZRDM is well described by a three state model based on the transcription cycle.

(A) Average FRAP of RNAP-GFP in EZRDM (blue) and with rifampicin treatment (pink). The dashed line represents the simulated FRAP curves using the best fit parameters from (B) and (C) for untreated and rifampicin treated cells respectively. (B) Model of transcription used to describe the FRAP data of RNAP-GFP in EZRDM with best fit parameters from simulations shown. (C) Simple two-state model used to describe the FRAP data of RNAP-GFPuv in cells treated with rifampicin with best fit parameters from simulations shown.

finding its promoter sequence, escaping the promoter to engage in transcription elongation, and finally completing transcription by dissociating, wherein the cycle can begin anew. To model this, we included three states: a freely diffusing state, a DNA-bound state, and an active elongation state with defined kinetic rates for promoter search, promoter escape, and transcription termination (Figure 2.4A).

We found this three-state model to describe our data well with 34% of RNAP freely diffusing, 26% of RNAP being DNA-bound, and 40% of RNAP undergoing active transcription. The pseudo-on rate was 0.89 s^{-1}, on par with the estimate of the initiation rate from *rrn* operons[58]. The transition rate from DNA-bound to freely diffusing was 1.1 s^{-1}, which likely represented a weighted average of the non-specific dissociation and abortive initiation rate. The promoter escape rate was determined to be 0.012 s^{-1}, which is in the correct order of magnitude from various biochemical studies (~10% of initiating RNAP could escape the promoter[6]), and the transcription termination rate, determined to be 0.0083 s^{-1}, corresponded to an average transcription elongation time of ~120 seconds, nearly identical to the previously measured average rRNA operon transcription time of ~130 seconds[58]. As most RNAP is engaged in rRNA synthesis under fast growth conditions, these results support our initial model.

In contrast, cells treated with rifampicin were well described by a two phase recovery - a rapid phase (~2s) likely representing free diffusion, and another much longer recovery (~500s) likely representing slow exchange of promoter-trapped RNAP. We developed a simple two-state model in which freely diffusing and DNA-bound RNAP could exchange with defined kinetic rates (Figure 2.4B). We used this model to simulate FRAP curves, and used least squares minimization to find kinetic parameters that minimized the squared difference between the simulated and experimental FRAP curves. For rifampicin-treated cells, we found this two-state model described the data well, with 39% of RNAP freely diffusing and 61% of RNAP DNA-bound. The kinetic rates between these two states were quite slow, with a k_{off} (bound → free) of 0.0011 s^{-1} and a k_{on} (free → bound) of 0.0017 s^{-1}, consistent with RNAP slowly exchanging on the DNA. These results corroborated the model for the rich growth condition, as the transition rates were several fold smaller than the rate we assigned to transcription termination. Ultimately, our results show that FRAP recovery in a WT context can be well explained by a simple three-state model, and that the resulting kinetic parameters match surprisingly well to prior *in vitro* biochemical studies.

B. SMT shows RNAP has three distinct diffusive states

Our FRAP data shows RNAP to be constantly undergoing the search-initiation-elongation-termination transcription cycle. However, this technique limits us to view the population average of all cells, and does not have the temporal resolution to accurately determine the search time. Additionally, we are unable to monitor the motion of RNAP as it undergoes the search process, disallowing us to probe the possible molecular mechanisms of promoter search. To overcome these limitations of FRAP, we performed single molecule tracking (SMT) experiments using a photoactivatible fluorescent protein PAmCherry C-terminally tagged to the β’ subunit of RNAP in live E. coli cells growing in EZRDM at RT. In this strain, the gene encoding PAmCherry is fused to the C-terminus of the rpoC gene encoding the β’ subunit of RNAP, and inserted at the native locus on the chromosome to replace the endogenous copy. We demonstrate that the rpoC-PAmCherry fusion protein is full-length and grew at a similar rate to the MG1655 parental strain (Figure 2.5). Thus, the rpoC-PAmCherry fusion protein functions fully as the sole cellular source of the β’ subunit of RNAP. In the text below, we use the term RNAP-PAmCherry to refer to our labeled RNAP subunit.

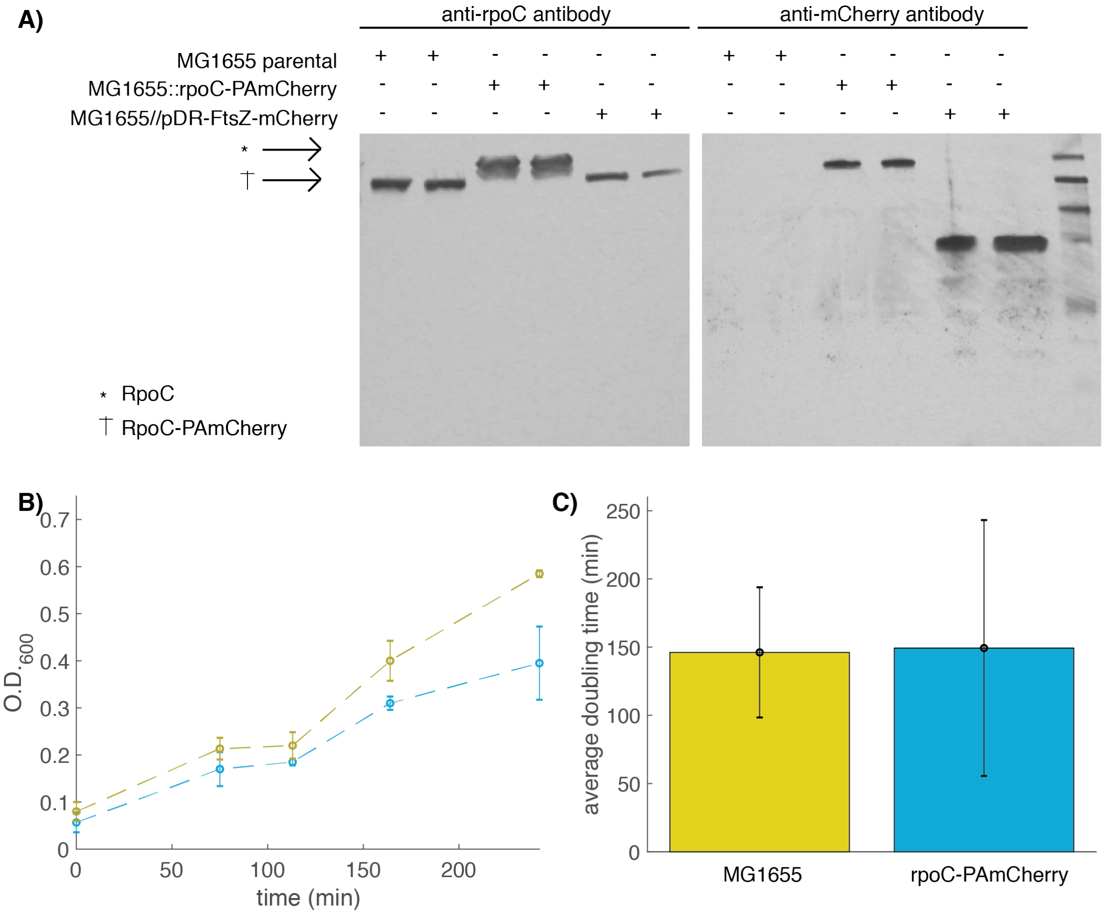


Figure 2.5: the rpoC-PAmCherry fusion can replace the endogenous rpoC for its function.

(A) Western blot on cell lysates from the MG1655 parental strain, the MG1655::rpoC-PAmCherry imaging strain, and a control MG1655//pCH-FtsZ-mCherry strain against the β' subunit of RNAP and mCherry. (B) Growth curve of one replicate of MG1655 and rpoC-PAmCherry strains. Error bars represent standard deviation from three technical replicates. (C) Quantification of the doubling time of MG1655 and rpoC-PAmCherry strains from five biological replicates. Error bars represent the standard deviation.

We stochastically activated individual RNAP-PAmCherry molecules in live cells and imaged their cellular positions with a frame rate of approximately 150 Hz (Δt = 6.74ms). At this frame rate, we were able to sample fast-moving molecules with an apparent diffusion coefficient, D_{app}, up to 3 $\mu m^2/s$. In addition, the uncertainty in determining the molecule's position, 30nm (Chapter 2, Section VI.B and Figure 2.6) through centroid fitting under our imaging conditions, limits the detection of

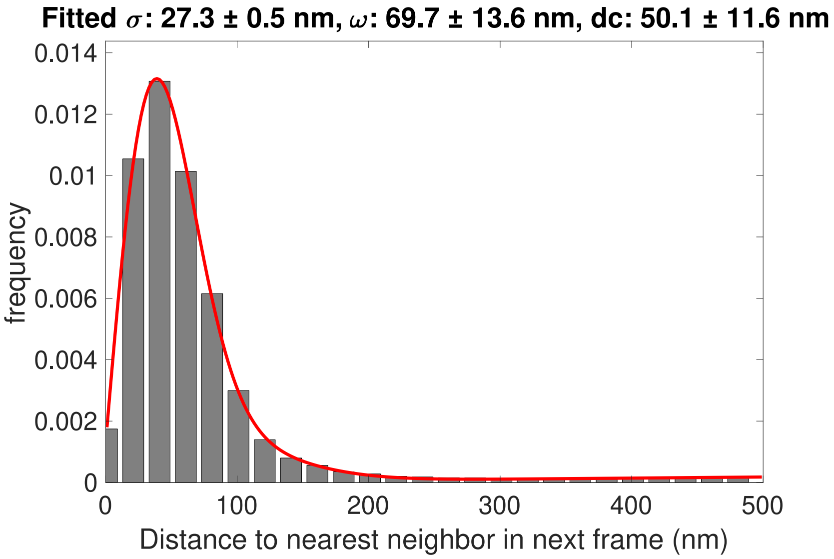


Figure 2.6: nearest neighbor fitting to determine the spatial resolution in our experimental setup.

In fixed rpoC-PAmCherry cells, individual molecules with repeat localizations were detected and the nearest neighbor distances were calculated and binned.

slow moving molecules to molecules with D_{app} >= 0.035 µm^2/s. Thus, RNAP-PAmCherry molecules diffusing within this range can be tracked and their trajectories mapped onto corresponding cell outlines with high confidence.

We collected a total of 63,182 individual RNAP-PAmCherry trajectories than a single frame from 353 cells, and plotted a few representative trajectories that displayed different diffusive properties (Figure 2.7). It is clear that individual RNAP-PAmCherry molecules diffused differently from one another, with some slow (magenta trajectory), some fast (blue trajectory), and some that switched between slow and fast diffusion (purple trajectory).

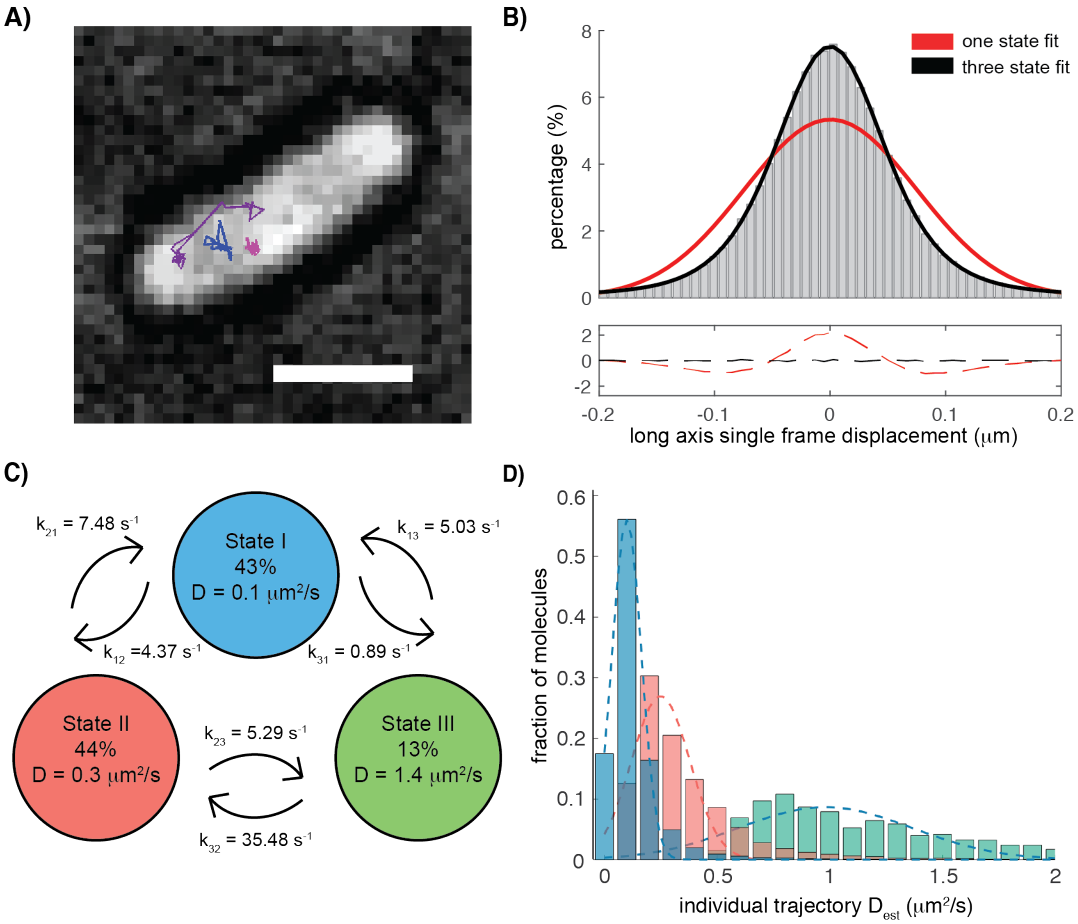


Figure 2.7: RNAP undergoing promoter search exhibit three distinct diffusive states that undergo state transitions

(A) Representative trajectories plotted on top of the brightfield image of the cell that display slow, confined motion (pink), fast motion (blue), and molecules that transition from slow to fast (purple). Scale bar represents one micron. (B) Histogram distribution of single frame displacements along the long axis of the cell (black dots). A single state model (blue) fits the data poorly, while a three state model (red) fits the data well. (C) HMM of RNAP-PAmCherry in fast-growing cells displays three diffusive states with their diffusion coefficients, state occupation percentages, and transition rates listed. (D) Individual trajectories were separated by their state assignments from the HMM in (C) and binned by their apparent diffusion coefficient. All three trajectories show distinct separation with lognormal averages (dashed lines) that mimic the values determined in the HMM from (C).

To investigate the diffusion behavior quantitatively, we plotted the distribution of the one-dimensional displacement *d1* along the long axis of the cell at single steps for all trajectories (Figure 2.7B). We found that this distribution

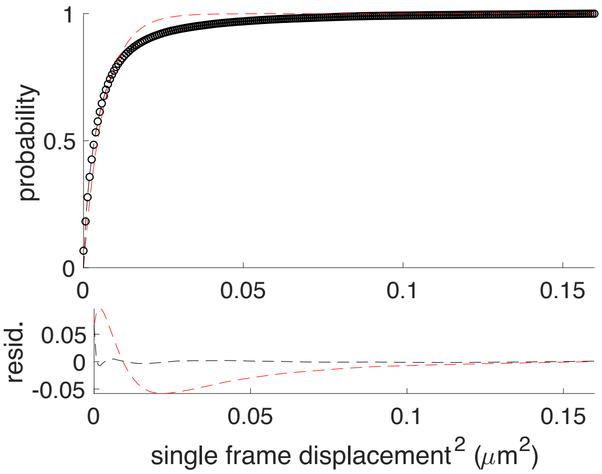


Figure 2.8: cumulative probability distribution of squared single frame displacements corroborates the presence of three diffusive states.

Single frame displacements were calculated and the cumulative probability of squared displacements less than or equal to a value *x* was constructed and fit to a model. A three state model (black) described the distribution much better than a single state model (red). Residuals of the two states are plotted below the CDF and highlight the poor fit of a single state model.

deviated from a single Gaussian function (Figure 2.7B, blue line), which is expected for a single population of molecules undergoing random diffusion, but instead was best fit to the sum of three Gaussian functions (Figure 2.4B, red line), suggesting the presence of three different diffusive states. Correspondingly, the percentages and diffusion coefficients of the three populations were found to be: 32.3% with 0.1 µm^2/s, 51.5% with 0.25 µm^2/s, and 16.4% with 1.31 µm^2/s. Additionally, the presence of three populations and their corresponding properties was supported by the analysis of the cumulative probability distribution of single-step displacements[48] (Figure 2.8).

C. Bayesian statistics provides a robust method for determining RNAP dynamics

The presence of three different diffusive states of RNAP suggests that at any given time, there are at least three populations of RNAP molecules engaged in different interactions with DNA, such as specifically bound, non-specifically bound, or randomly diffusing. These states cannot be static, as RNAP must non-specifically bind and unbind from the DNA many times during promoter search. This suggests a single RNAP molecule could transition from one state to another as it interacts with the DNA during promoter search, until it binds specifically to the DNA to initiate transcription. Traditional methods of finding the kinetics of state switching rely on complicated modeling, or separating trajectories into separate diffusive states using a user-defined threshold. This type of thresholding is problematic for molecules with smaller diffusion coefficients, or whose diffusive states are close in value, due to extensive overlap between the single frame displacement distributions, which gives an unacceptably high error rate in state assignments.

To overcome these limitations, we utilized Bayesian statistical algorithms from the software vbSPT to form a hidden Markov model (HMM) of the single molecule tracking data[50]. This method is able to combine thousands of SMT data together to uncover several properties: the optimal number of diffusive states that best describes the data, the diffusion coefficients of those states, and the likelihood that an individual molecule will switch states, aka the transition probability. The resulting HMM of our tracking data of RNAP-PAmCherry separated RNAP into three diffusive states with population percentages and coefficients of: 43% with 0.1 $\mu m^2/s$, 44% with 0.3 $\mu m^2/s$, and 13% with 1.4 $\mu m^2/s$, respectively (Figure 2.7C). Although State 1 and State 2 have diffusion coefficients close in value, they show distinct separation, as with our prior analyses. Separating the displacements of individual trajectories based on their state assignments shows these diffusive states are well separated from one another (Figure 2.7D), and closely resembles the results from the single frame displacement distribution fitting (Figure 2.7B) and cumulative displacement probability distribution fitting (Figure 2.8).

A distinct advantage of the HMM over the previous analyses is the kinetic information it provides. Similar to the diffusion coefficient calculation, the software utilizes a maximum likelihood algorithm to calculate the probability that a given molecule will transition from one state to another. These transition probabilities can be converted to kinetic rates through a straightforward mathematical derivation[49]. Using these kinetic rates, the time that RNAP stays in a given state, or dwell time, can be calculated. The dwell times were calculated to be 200ms, 90ms, and 30ms for states 1, 2, and 3, respectively. The diffusion coefficients, occupation probabilities, and transition probabilities are tabulated in Table 2.1.

We theorized that each of the diffusive states represented a particular action of RNAP: state 1 could be RNAP that is bound to the DNA, state 2 could be an RNAP that is undergoing rapid promoter search within the nucleoid, and state 3 could be freely diffusing RNAP. If these hypotheses are true, then a dwell time of 200ms for state 1, predicted to be DNA-bound RNAP, is an insufficient time for RNAP to undergo the many steps of transcription, which we determined to be ~ 120s (Figure 2.4). Thus, it is likely that this state represents RNAP that is

	state 1			state 2			state 3			
Condition	D (µm²/s)	state %	τ (ms)	D (µm²/s)	state %	τ (ms)	D (µm²/s)	state %	τ (ms)	N (# displ.)
EZRDM	0.10 ± 0.002	43.5 ± 1.2	198 ± 20	0.30 ± 0.007	43.5 ± 1.1	86 ± 6	1.44 ± 0.03	13 ± 0.4	27 ± 1	63,182
I1309A	0.10 ± 0.003	50 ± 1.5	268 ± 75	0.38 ± 0.007	50 ± 1.5	56 ± 6				13,473
+ rif	0.17 ± 0.004	37 ± 0.8	117 ± 15	0.83 ± 0.008	63 ± 0.8	82 ± 7				40,144
+ chlor	0.09 ± 0.001	48 ± 0.7	278 ± 29	0.39 ± 0.003	52 ± 0.7	62 ± 3				44,168
M9	0.11 ± 0.003	44 ± 1.3	196 ± 42	0.38 ± 0.005	56 ± 1.3	91 ± 10				21,500
AsiA o/e	0.11 ± 0002	44 ± 1	229 ± 35	0.39 ± 0.004	56 ± 1	87 ± 7				33,965

Table 2.1: tabulation of HMM values of RNAP-PAmCherry in all imaged conditions.

Diffusion coefficients, state percentages, and dwell times (τ) are listed for each diffusive state in each imaged condition. Error is from bootstrapping and represents the s.e.m.

non-specifically bound to the DNA. However, it must be noted that the HMM and maximum likelihood methods cannot distinguish between states that have identical diffusion coefficients, but different dwell times, as would be the case for RNAP that is non-specifically versus specifically bound. However, given the short dwell time which is three orders of magnitude less than the calculated transcription rate from our FRAP experiments, we are confident that state 1 more likely represents non-specifically bound RNAP. Next, we aim to provide evidence for our hypothesized state assignments and parse out a complete dynamical model of RNAP *in vivo*.

D. Assignment of state 1 as DNA-bound RNAP

Intuitively, it would follow that state 1 of the HMM of RNAP, which represents the slowest state of RNAP, would be an RNAP that is bound to the DNA. Additionally, the value we determined for state 1, 0.1 µm^2/s, is identical to the reported values for the diffusion coefficient of DNA in *E. coli*, as well as the values we reported for the nucleoid-associated protein HU (Chapter 3). The size

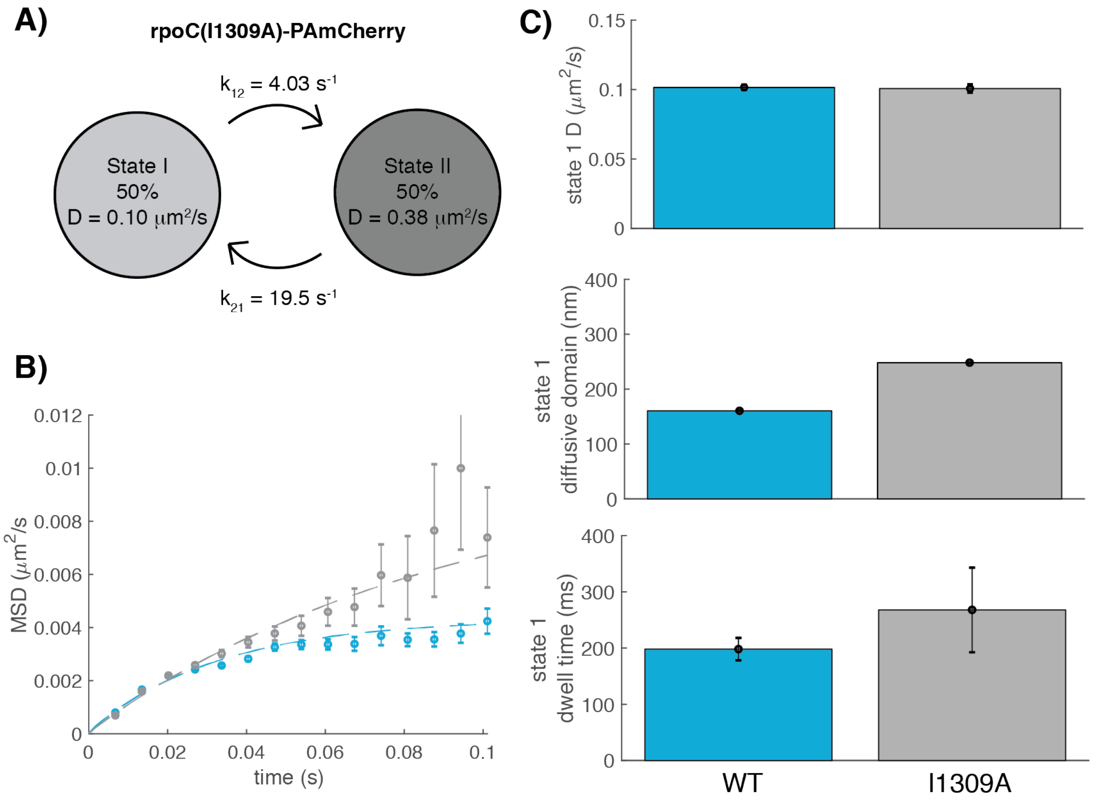


Figure 2.9: state 1 RNAP-PAmCherry likely represents RNAP non-specifically bound to DNA.

(A) HMM of RNAP(I1309A)-PAmCherry grown in EZRDM exhibits two diffusive states. Diffusion coefficients, occupation percentages, and transition rates are listed. (B) MSD of WT State 1 RNAP-PAmCherry (blue) and State 1 RNAP(I1309A)-PAmCherry (gray). The deviation from linear behavior is indicative of confinement. (C) Comparison of the diffusion coefficient, diffusion domain size, and dwell time values for WT State 1 RNAP-PAmCherry and State 1 RNAP(I1309A)-PAmCherry. Values are μ ± s.e.m. (from bootstrapping).

discrepancy between RNAP and HU is almost two orders of magnitude, suggesting their diffusion coefficients should differ by several-fold. However, the nearly identical diffusion coefficients suggest these molecules are bound by something much larger than the two, such as DNA. DNA is a large, relatively immobile molecule, meaning that a molecule bound to DNA would have the diffusion coefficient of the DNA.

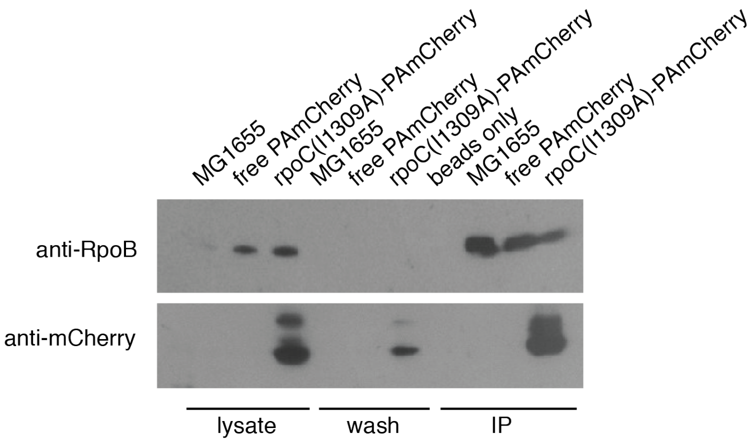


Figure 2.10: RpoC(I1309A)-PAmCherry is able to be incorporated into the core RNAP enzyme.

Cell lysates from MG1655, MG1655//pCH-PAmCherry, and MG1655//pCH-rpoC(I1309A)-PAmCherry were incubated with beads containing anti-RpoC antibody and detected on protein blots using anti-RpoB and anti-mCherry antibodies. Lanes 1-3 are the cell lysate, lanes 4-6 is the wash step, lane 7 is the beads-only control, and lanes 8-10 is the immunoprecipitation elute. All three strains were able to detect RpoB, as RpoC and RpoB both incorporate into the holoenzyme. The lack of signal in the free PAmCherry lane in the IP step when using the anti-mCherry antibody indicates no cross-reaction between PAmCherry and the anti-RpoC antibody.

To help corroborate this supposition, we examined the mean-squared displacement (MSD) plot of state 1 RNAP-PAmCherry, calculating along the x-axis. Normal Brownian motion has a linear relationship between the MSD and time proportional to its diffusion coefficient. However, if a particle is confined to a local area, there is an asymptotic relationship to the MSD as time increases. Indeed, we observed that the State 1 RNAP-PAmCherry MSD deviated from linear behavior (Figure 2.9B). Using an equation that assumes confinement within a box of length L, the confinement area, or diffusive domain size, can be calculated. For state 1 RNAP-PAmCherry, the diffusive domain size is ~ 160 nm (Figure 2.9C), similar to our previous reported domain size for DNA-bound HU and a chromosomal DNA locus (Chapter 3, ~ 230nm and 200nm respectively).

Together with the diffusion coefficient, this data supports the hypothesis that state 1 RNAP-PAmCherry is bound to the DNA.

Above, we theorized that this RNAP-PAmCherry state 1 is non-specifically bound to the DNA. To support this theory, we imaged a mutant of the β' subunit of RNAP, β'(I1309A), which has been shown *in vitro* to have significantly poor ability to form an open complex on DNA, meaning it should only be able to bind DNA non-specifically. We expressed this mutant, C-terminally tagged with PAmCherry, on an inducible plasmid and show that it can still be incorporated into the holoenzyme using co-immunoprecipitation (Figure 2.10). We performed similar SMT experiments on cells grown in EZRDM media and collected 13,473 trajectories from 275 cells. Using the same HMM analysis with identical algorithm parameters, we find there are two diffusive states: one with diffusion coefficient of 0.1 $\mu m^2/s$, and one with diffusion coefficient of 0.32 $\mu m^2/s$. The state 1 diffusion coefficient in the I1309A mutant is almost identical to WT state 1 RNAP-PAmCherry, has similar diffusive domain size, and a similar dwell time of 260ms (Figure 2.9C). As this mutant can only form a closed complex on the DNA, this supports the hypothesis that the majority of what we view in our SMT of RNAP is non-specifically bound RNAP.

E. Assignment of state 2 as RNAP undergoing promoter search in the nucleoid

We theorized that state 2 RNAP, corresponding to a diffusion coefficient of 0.3 $\mu m^2/s$, was RNAP undergoing rapid promoter search within the dense nucleoid of the cell. When we looked at the MSD plot of state 2 RNAP-

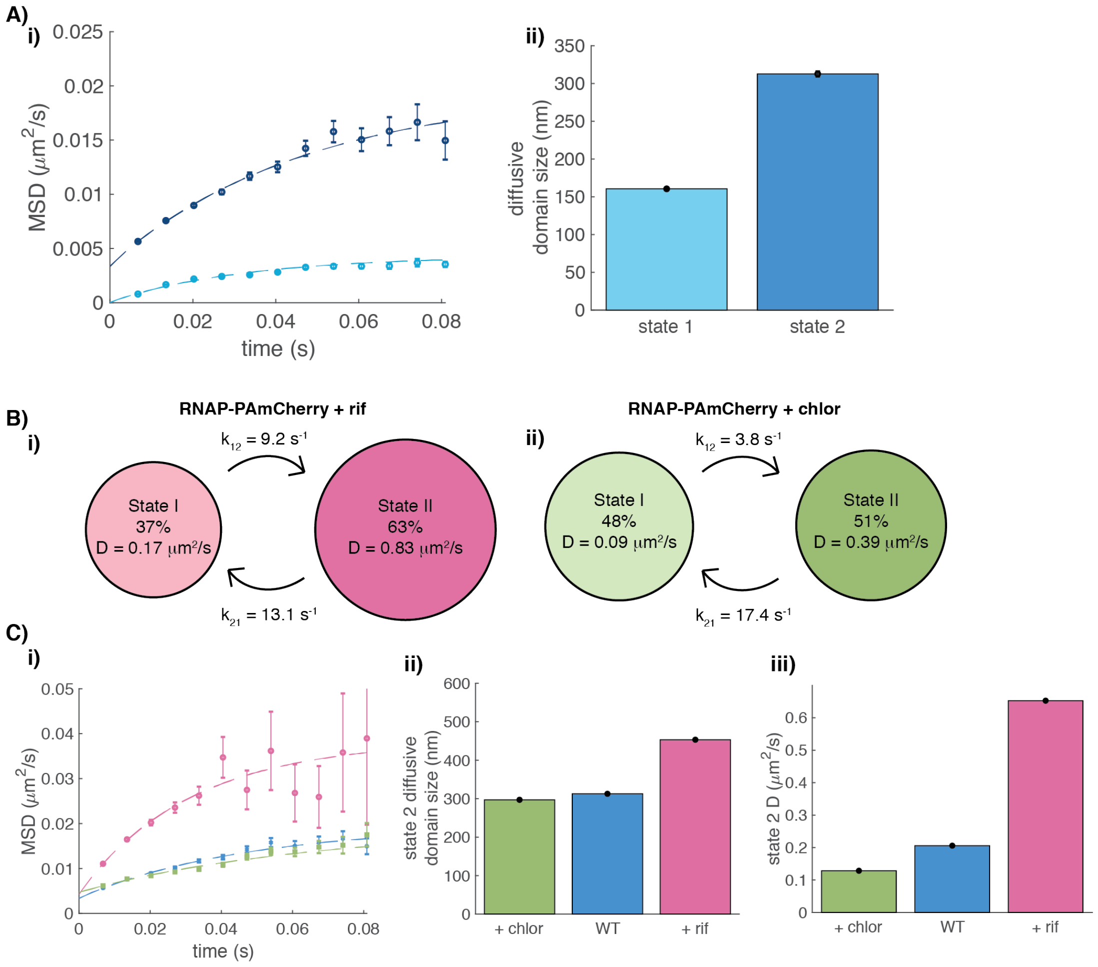


Figure 2.11: state 2 RNAP likely represents RNAP diffusing within the nucleoid "mesh".

(A) MSD (i) of state 1 (light blue) and state 2 (dark blue) RNAP-PAmCherry demonstrating a larger confinement area for state 2 over state 1 as indicated by the values (ii) of the diffusive domain size to the right. Values are μ ± s.e.m. (B) HMM of RNAP-PAmCherry cells treated with rifampicin (i) and chloramphenicol (ii). Diffusion coefficients, occupation percentages, and kinetic rates are listed. (C) Comparison of the MSD (i) of state 2 RNAP-PAmCherry in chloramphenicol-treated (green), untreated (blue), and rifampicin-treated (red) cells with diffusive domain values (ii) and diffusion coefficients (iii) values compared. Values are μ ± s.e.m.

PAmCherry, we found that the diffusive domain of this state was relatively small compared to the size of the cell, but larger than the diffusive domain of state 1 (Figure 2.11A). We theorized that this diffusive domain size was due to RNAP being confined to a particular region of the chromosome due to its relative size

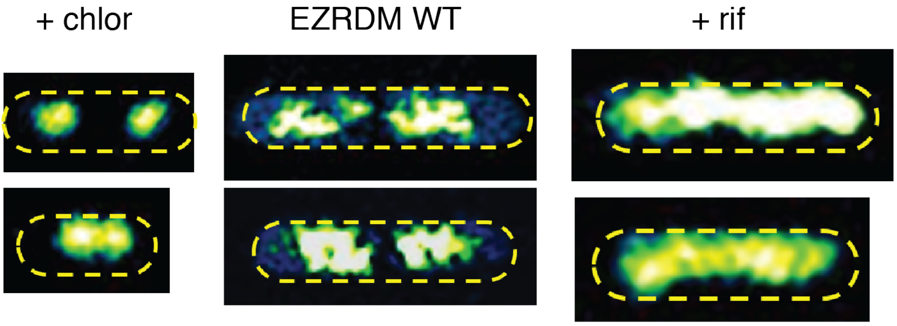


Figure 2.12: effects of chloramphenicol and rifampicin drug treatments on nucleoid size.

Chloramphenicol-treated (left), untreated (middle), and rifampicin-treated (right) cells were stained with 10nM Hoechst 33342 dye and imaged using SIM microscopy. Images are false colored to maximize the contrast and approximate cell outlines are in yellow dashed lines.

and the local density of DNA. Thus, we predict the diffusive domain size to correlate to the average "pore" size of the dense nucleoid matrix. To test this hypothesis, we performed perturbations to the cell using rifampicin and chloramphenicol, both of which have known effects on the compaction state of the nucleoid. First, we performed SMT experiments on RNAP-PAmCherry in cells treated with rifampicin (200 µg/mL, 15 min). Under rifampicin treatment, the nucleoid expands to fill much of the cytoplasmic volume[59] (Figure 2.12). Performing the same HMM analysis as before to extract diffusive parameters, we find that the diffusion coefficient of state 2 RNAP increases from 0.3 µm^2/s to 0.83 µm^2/s (Figure 2.11Bi), and that the diffusive domain increases from 0.3 microns in diameter to 0.5 microns (Figure 2.11C). This is consistent with our hypothesis that RNAP is diffusing in the nucleoid, and is confined to a local space dictated by the condensation of the nucleoid. Despite the moderate change in the diffusion coefficient and diffusive domain size of state 2, the diffusion coefficient and diffusive domain size of state 1 RNAP-PAmCherry for both untreated and rifampicin-treated cells only sees a modest change.

To further probe our hypothesis, we also performed SMT experiments on RNAP-PAmCherry cells treated with chloramphenicol. We show that the nucleoid does indeed show smaller compaction relative to untreated cells, as previously reported[59] (Figure 2.12). According to our hypothesis, this should decrease the “pore” size of the nucleoid, and we should see a decrease in both the diffusion coefficient of state 2 RNAP-PAmCherry and its diffusive domain size. We observed a modest decrease in diffusion coefficient of state 2 RNAP, and a slight change in the diffusive domain size, decreasing from ~312 nm in untreated cells to ~ 297 nm in chloramphenicol-treated cells. Using the confinement-corrected diffusion coefficient from our MSD fit, we found that the diffusion coefficient of state 2 RNAP-PAmCherry in chloramphenicol-treated cells was smaller than that of untreated cells. Overall, these data support our hypothesis that state 2 RNAP is diffusing within the nucleoid, and experiences significant confinement effects due to the organization of the nucleoid.

F. Assignment of state 3 to freely diffusing RNAP

It intuitively follows that if state 1 is RNAP that is bound to the DNA, and that state 2 is RNAP that is diffusing within the nucleoid, then state 3, with the fastest diffusion coefficient of the three states at 1.44 $\mu m^2/s$, would be RNAP that

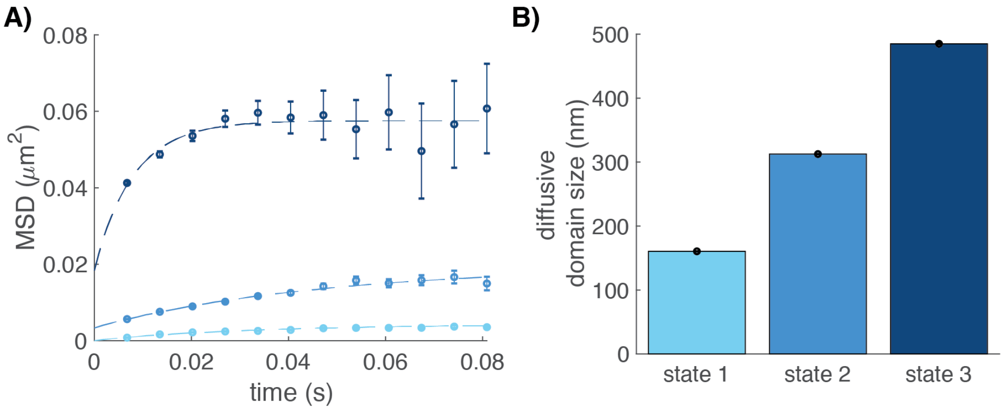


Figure 2.13: state 3 RNAP-PAmCherry experienced much larger confinement than state 1 and 2, consistent with a freely diffusing state.

(A) Comparison of the MSD of state 1 RNAP-PAmCherry (light blue), state 2 RNAP-PAmCherry (medium blue), and state 3 RNAP-PAmCherry (dark blue) in EZRDM growth media. (B) Comparison of the diffusive domain sizes of state 1, state 2, and state 3 RNAP-PAmCherry. Values are μ ± s.e.m. from bootstrapping.

is freely diffusing in the cell. Indeed, when we plotted the MSD of state 3 RNAP-PAmCherry, we found that its diffusive domain was significantly larger than either state 1 or 2 RNAP-PAmCherry (Figure 2.13). Similarly, using experimentally derived values for the size of RNAP and the viscosity of the cell, we find that according to the Stokes-Einstein equation, the diffusion coefficient of RNAP should be around 1.5 μm^2/s. A previous study of RNAP diffusive motion found that in replication-stalled cells, where the poles were DNA-free, RNAP had a distribution of diffusion coefficients with a mean of ~ 2 μm^2/s[60]. Together, this supports a state assignment for state 3 RNAP of freely diffusing.

G. RNAP dynamics deviate from expected steady state behavior

Now that we have plausible state assignments for each diffusion coefficient of RNAP-PAmCherry, with state 1 being DNA-bound RNAP, state 2 being RNAP diffusing in the nucleoid, and state 3 being freely diffusing RNAP, we can now better understand the dynamics of RNAP movement as it undergoes promoter search. First, we notice that the dwell time of RNAP is longest for state 1 RNAP, at 200ms, and shortest for state 3, at 30ms. These kinetics suggest a preference for RNAP to be bound to the DNA. In our HMM, the occupation percentages of the diffusive states are calculated independently of the kinetic rate values, meaning it is possible for the observed occupation percentages and predicted occupation percentages based on the kinetic rates to differ. Indeed, we find that the observed occupation percentages deviate from the predicted values (Figure 2.14).

Perhaps unsurprisingly, the nucleoid compaction state seems to have an effect on the search dynamics. In rifampicin treatment, where the nucleoid is more expanded, we observe the dwell time of state 1 RNAP reduces from ~200ms to ~120ms. In chloramphenicol treatment, the opposite is true where the dwell time increases to ~ 280ms. Similarly to untreated cells, the observed percentage of RNAP in chloramphenicol or rifampicin treated cells does not match the values predicted by the kinetic rates (Figure 2.14). This puzzling observation suggested to us that RNAP kinetics might not be in steady state. Prior theoretical work[44,61,62] and some experimental work[26,43] have found that crowding can cause biochemical reactions to deviate from mass action (where the assumption is a well-mixed system) of exponential rate laws to rate laws that have a power law dependence. Long-tailed distributions would over- or underestimate the apparent rate transitions, leading to a discrepancy in the

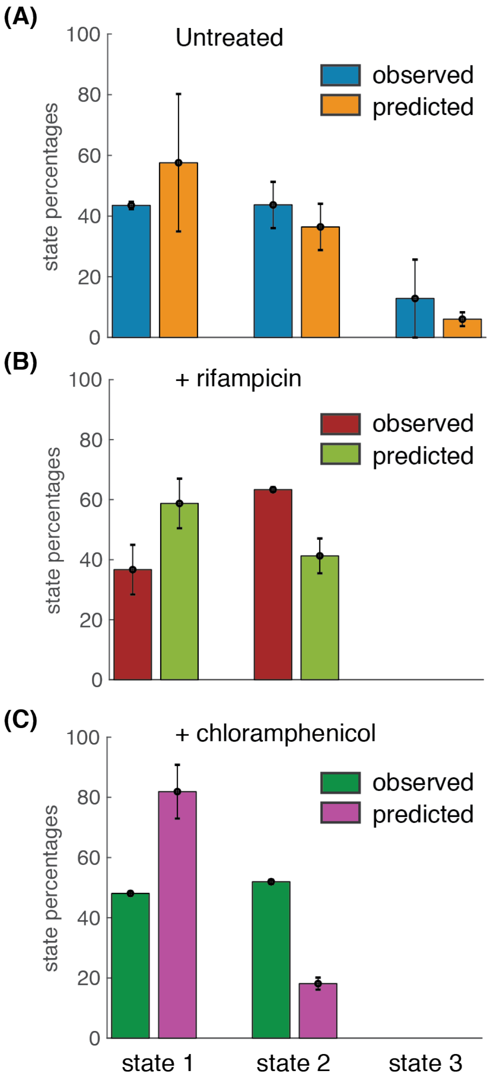


Figure 2.14: observed state occupation percentages differ from the predicted values calculated from the independently-determined kinetic rates.

Comparison of the observed state occupation percentages and the predicted values calculated from the independently-determined HMM transition rates for (A) untreated RNAP-PAmCherry in EZRDM, (B) RNAP-PAmCherry with rifampicin treatment, and (C) RNAP-PAmCherry with chloramphenicol treatment. Note that in each condition, these two values differ substantially, indicating possible deviation from steady state behavior.

predicted occupation percentages and the observed occupation percentages. However, more work is needed to explore this idea.

H. Transcription kinetics are sensitive to the metabolic state

Thus far we have mainly focused on RNAP in rich growth media, where the vast majority of transcription is dedicated to ribosomal gene synthesis. These *rrn* operons are unique in that they have a very strong promoter and are thus predicted to contain a higher than average number of RNAP per gene. We wondered if shifting transcription to mRNA synthesis would lead to changes in the transcription kinetics. To probe this, we performed FRAP experiments on RNAP-GFPuv cells grown in minimal M9 media, or with overexpression of the T4 phage anti-σ70 factor AsiA, which sequesters σ70 (the housekeeping σ factor) and prevents transcription from these genes[63].

For RNAP-GFPuv in M9 media, we found that as with the fast growth condition, the recovery dynamics were best described by the three state model based on the transcription initiation-elongation-termination cycle (Figure 2.15). Interestingly, we do observe differences in the kinetics. First, the k_{on} and k_{off} rates decreased and increased respectively compared to the fast growth condition, suggesting that RNAP-PAmCherry search strategy is changed. Second, the promoter escape and transcription termination rates slightly increased. The increase in promoter escape rate may be due to the lowered promoter strength of most mRNA genes compared to the *rrn* operon. Overall these respective rate changes lead to a substantial increase in freely diffusing RNAP from 34% to 42% while maintaining a similar relative population of elongating RNAP (~40%).

For RNAP-GFP with AsiA overexpression, we found the three-state model to again describe the data well (Figure 2.15). We again observe altered

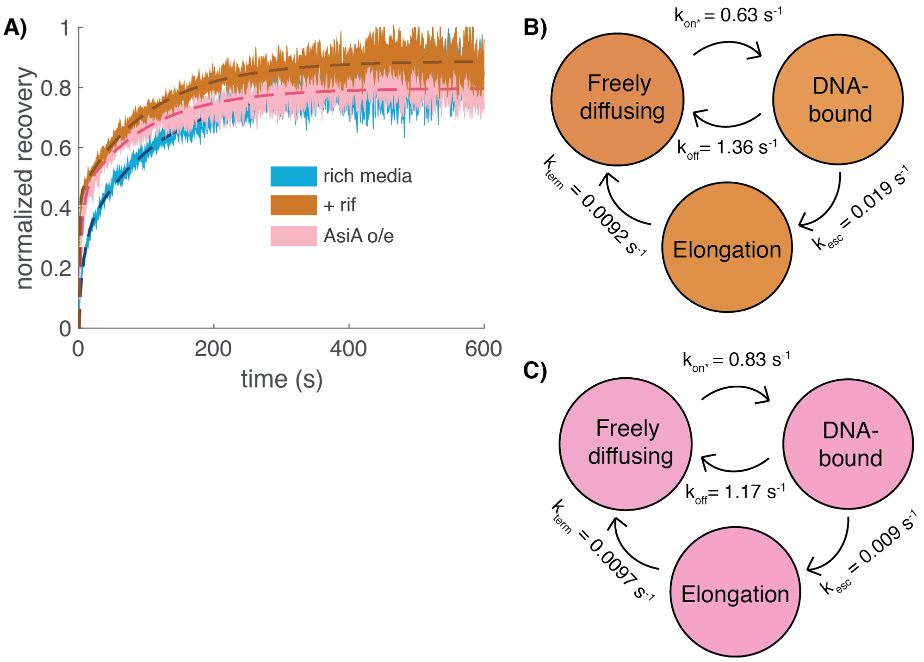


Figure 2.15: FRAP data demonstrates altered transcription kinetics when the metabolic state changes.

(A) Average normalized recovery curves for the M9 condition (orange) and AsiA overexpression condition (pink) compared to the WT condition (blue). Dark smooth lines are the simulated FRAP curves based on the best fit models in (B) and (C). (A) The best fit model for RNAP-GFPuv in M9 media with best fit model parameters listed. (B) The best fit model for RNAP-GFPuv with AsiA overexpression with best fit model parameters listed.

transcription kinetics. Similarly to the M9 condition, RNAP-GFPuv with AsiA overexpression had a lowered k_{on} and an increased k_{off}. Unlike the M9 condition, the promoter escape rate for the AsiA overexpression condition was reduced, while the transcription termination rate was increased. This likely reflects the relative changes in gene expression between these two conditions. In the M9 condition, cells must utilize their σ70 housekeeping genes to survive, while in the AsiA condition, this transcription was abolished. Overall, these rate changes lead to an increase in freely diffusing RNAP, from 34% to 43%, and a decrease in transcribing RNAP from 40% to 27%. These data together suggest that the

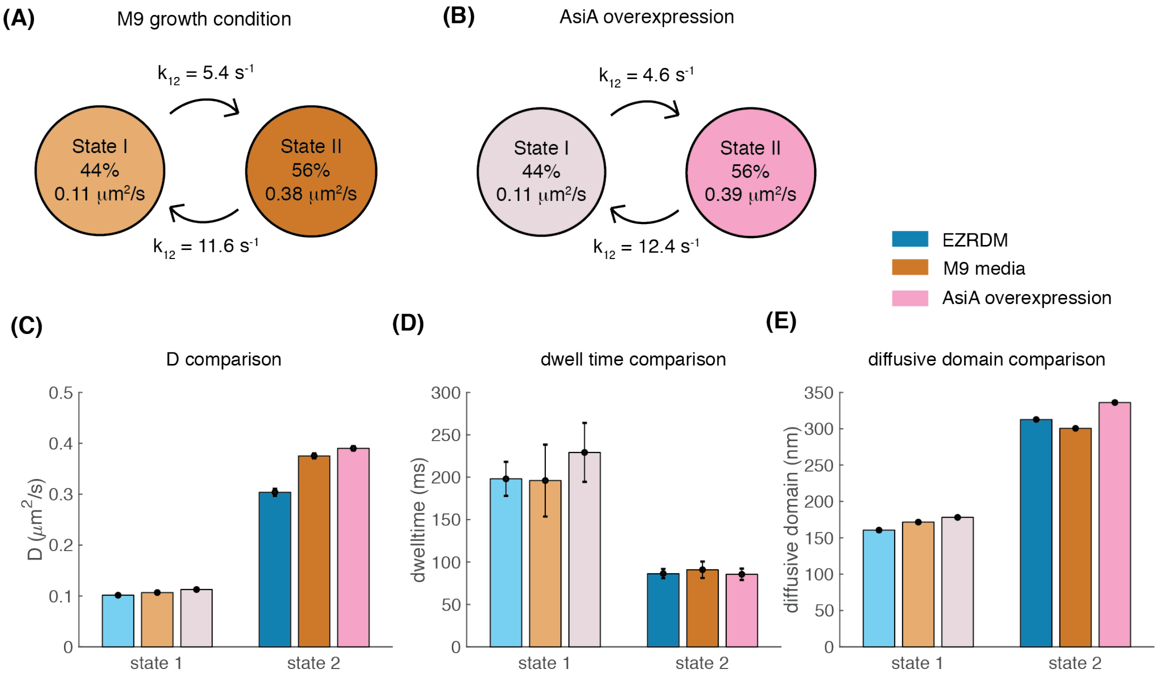


Figure 2.16: altered metabolic state had negligible effect on RNAP search dynamics.

HMMs for (A) RNAP-PAmCherry in M9 growth media and (B) RNAP-PAmCherry with AsiA overexpression are depicted with their respective diffusion coefficients, state occupation percentages, and transition rates. Comparison of (C) the diffusion coefficients, (D) the state dwell time, and (E) the diffusive domain size for RNAP-PAmCherry in rich growth media (blue), M9 growth media (orange), and AsiA overexpression (pink) are shown with error bars representing the standard deviation as determined by bootstrapping.

metabolic state- and thus the gene expression output- influence the transcription cycle kinetics.

I. Metabolic state has negligible effect on promoter search dynamics

Our FRAP results suggested that RNAP-GFP has a more difficult time sampling non-specific DNA in the M9 media and AsiA overexpression conditions as compared to the fast growth condition. To test the validity of this, we performed SMT on RNAP-PAmCherry in M9 media and under AsiA overexpression. Our HMMs of these two conditions demonstrated the presence

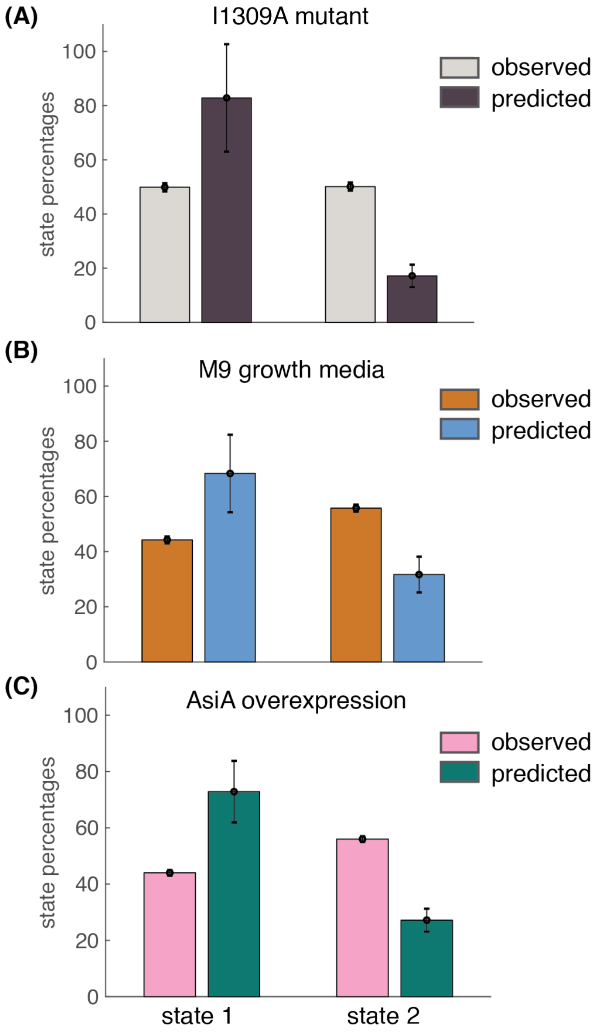


Figure 2.17: altered transcription states also displayed differences between observed and predicted state occupation percentages.

Comparison of the observed state occupation percentages compared to the predicted values calculated from the independently-determined HMM transition rates for (A) RNAP(I1309A)-PAmCherry mutant, (B) RNAP-PAmCherry grown in minimal M9 media, and (C) RNAP-PAmCherry with AsiA overexpression. Error bars represent standard deviations.

of two diffusive states with diffusion coefficients that closely mimicked the diffusion coefficients of states 1 and 2 RNAP-PAmCherry in EZRDM. Moreover, the dwell times and diffusive domain sizes of states 1 and 2 in all three conditions were comparable. Together, this would suggest that the underlying search process remains the same in all three conditions, i.e. that the search process is insensitive to gene expression output. In our FRAP model, the DNA-bound state represents RNAP bound to both non-specific and promoter sequences, thus the off rate is a combination of the failed open complex formation and non-specific binding dissociation. Thus, it is possible to have identical non-specific binding dissociation rates while having different promoter dissociation rates.

Like the other conditions, the M9 and AsiA overexpression conditions (as well as the I1309A mutant RNAP) had observed state occupation percentages that differed from the predicted state occupation percentages calculated from the kinetic rates, again suggesting a deviation from steady state dynamics (Figure 2.17). This deviation from steady state dynamics could also explain the differences between the FRAP and SMT data. If the dissociation and/or association rates follow a power law distribution as opposed to an exponential distribution, the rates may again be over- or underestimated. This also may explain why the three state model has slightly worse fitting in the M9 and AsiA conditions compared to the EZRDM condition.

III. Discussion

Using FRAP, we first monitored the transcription cycle of RNAP. We found that the FRAP dynamics were best fit to a three state model based on the transcription initiation-elongation-termination cycle: freely diffusing RNAP can bind to DNA; if the DNA is non-specific it can dissociate and if the DNA is a promoter sequence then it can either dissociate or escape the promoter to undergo transcription elongation. Transcription elongation must terminate, and upon termination becomes freely diffusing again. Despite the plethora of literature on the effects of crowding on biochemical rates, the values we obtained from this model closely matched the values found in the literature via biochemical studies.

When we altered the gene expression profile by growing the cells in minimal M9 media, where mRNA synthesis is dominant, or under AsiA overexpression, where housekeeping gene synthesis is abolished, we observed FRAP recovery dynamics that could still be described by the three state model, but with altered kinetic rates. Generally, RNAP in rich growth media had a faster on and off rate from freely diffusing to DNA-bound than in the other two conditions, leading to fewer freely diffusing RNAP in EZRDM comparatively. We found that in minimal M9 media, the promoter escape rate increased compared to the EZRDM condition. As the majority of transcription is dedicated to the rrn operons in rich growth media, and that the promoters of *rrn* operons are stronger than most genes, we predict this reflects the relative differences in the average gene promoter strength, considering promoter escape is considered a rate limiting step in transcription. Interestingly, in the AsiA overexpression condition, the promoter escape rate decreased. This may be to prevent stress genes from being unnecessarily expressed under basal conditions. Overall, the information gleaned from these experiments will likely be useful to experimenters interested in gene expression. Given the differences in parameters between metabolic conditions, this methodology could be useful for determining the relative kinetics of various transcriptional profiles.

Through our single molecule tracking of RNA polymerase in live *E. coli* cells under a variety of transcriptional and genetic perturbations, we find RNA polymerase has three diffusive modes: (1) DNA-bound or DNA-associated, (2) diffusing in the dense nucleoid matrix, and (3) freely diffusing in the cell. We hypothesize that state 1 RNAP is truly bound to the DNA, and not undergoing facilitated diffusion. Multiple *in vitro* studies have found that unlike other DNA-binding proteins, such as the transcriptional repressor protein LacI, RNAP does not appear to undergo any kind of one-dimensional sliding but rather finds its promoter through purely random collisions[45,46]. In the study by Friedman and colleagues, fluorescently labeled DNA with a single strong promoter and variable upstream DNA lengths were adhered to slides and the search time was monitored as a function of the upstream DNA length. Even for the longest upstream DNA length, they found no change in the association rate of RNAP to the promoter, strongly implying RNAP locates the promoter through random collisions. Nonetheless, it is possible that state 1 RNAP is merely associated with the DNA, and not always non-specifically bound to the DNA. For example, Cravens and colleagues determined that human DNA glycosylase undergoes DNA-associated hopping events over extremely short distances-10s of base pairs- that would not be detectable under our imaging conditions[38]. However, given the prior results, we theorize that RNAP likely does not undergo facilitated diffusion, or at least does not slide long distances as other reported transcription factors.

Additionally, we find that RNA polymerase strongly favors the nucleoid environment, as the vast majority of RNA polymerases- 87 to 100%- are in states 1 and 2, depending on the transcriptional condition. This combined with the significantly small diffusive domain sizes of these two states leads us to speculate that RNA polymerase is retained mostly in these two states through crowding of the nucleoid and the pore size of the DNA "meshwork". Interestingly, while RNAP exhibited altered diffusive and confinement behaviors under different nucleoid sizes, another DNA-binding protein, the nucleoid-associated protein HU, exhibited almost identical diffusive and confinement behavior. Prior work has shown that different transcription factors can exhibit different search strategies[42]. The authors speculate that the binding partners of these transcription factors may dictate their search strategies. Here, given that both RNAP and HU primarily bind dsDNA, it is unclear what physical properties dictate their different diffusive behaviors. One possibility is their respective differences in size; because RNAP is much larger than HU, it may be much more greatly influenced by changes in nucleoid structure, similar to how molecules move within a gel. Interestingly, despite their differential changes upon nucleoid expansion or contraction, the diffusion coefficient of the "searching" RNAP and the diffusing HU are extraordinarily similar, despite the order of magnitude difference in their molecular weights. Both proteins localize almost exclusively to the nucleoid, indicating that some property of DNA-binding that keeps them nucleoid-localized influences their diffusive behavior within the nucleoid. Consistent with this, many other nucleoid-localized proteins in *E. coli* exhibit nearly identical diffusion coefficients within the nucleoid, including the transcription factor LacI[41], and the DNA repair enzyme UvrA[64].

Intriguingly, in all imaging conditions we saw significantly different observed state occupation percentages compared to the predicted state occupation percentages calculated from the independently determined kinetic rates. To verify that this observed effect is not due to a limitation of the software, we simulated various dynamical models over a range of diffusion coefficients and transition rates and found that vbSPT could accurately determine the ground truth parameters, even for short trajectories. Thus, we hypothesize that RNAP deviates from steady state dynamics, or that the dynamics are not well explained by the laws of mass action. Given the plethora of theoretical studies demonstrating biochemical rates deviating from typical mass action laws[36,37,44,61,65,66], as well as some experimental evidence of this behavior[35,43,67], this hypothesis seems quite plausible. However, more study will be needed to better understand the molecular mechanisms behind the observed dynamical behavior.

IV. Materials and Methods

A. Cell length comparison between rpoC-GFPuv strains and MG1655 parental strain

To assess the growth, we determined the cell length distribution of the MG1655 parental strain and compared it to the distribution for MG1655::rpoC-GFPuv in the rich growth media, minimal M9 media, arabinose-only control and

AsiA overexpression conditions. Cell lengths were determined from brightfield images by manually determining the length between the two cell poles.

B. FRAP data collection and analysis

For FRAP experiments, cells were inoculated from freshly streaked LB plates into 2mL of EZRDM or M9 media and let grown overnight (~16 hours) at 24°C with shaking. Saturated cultures were diluted 1 to 200 into fresh 2mL of their respective media and let grown for several hours to reach mid-log phase growth (O.D.600 ~ 0.4). For rifampicin treatment, rifampicin stock at 50mg/mL was added such that the final concentration was 200 µg/mL and incubated with shaking at 24°C for 15 minutes. For AsiA overexpression, the strain containing the pDR-pBAD-AsiA plasmid was spun down at OD600 ~ 0.25 and resuspended in EZRDM media with glycerol as the main carbon source with 0.2% arabinose added to induce AsiA expression. Cells were induced for 2 hours at 24°C with shaking. Cells in all conditions were washed once then spun and resuspended in 1/20th volume to concentrate the cells.

Harvested cells were pipetted onto agarose pads and sandwiched between the agarose pad and a #1 coverslip as previously described[68]. Immobilized cells were imaged on an Olympus IX71 inverted microscope with a 100x oil objective (NA =1.45). Photons from cells were collected with an Andor EMCCD camera using MetaMorph imaging software (Molecular Devices). Fluorescence from cells was obtained using solid-state lasers at 488nm wavelength (Coherent). Twenty frames at 100ms exposure per frame were taken before cells were bleached. Cells were bleached along the quarter cell by focusing the laser light onto the imaging plane. The ratio of the confocal beam and epifluoresence beam intensities was 100 to 1. Usually, and area encompassing ~20% of the total cell area was bleached using this set-up. Fluorescence recovery was monitored by acquiring 100ms fluorescence images every second for 600 frames (10 minutes).

To account for photobleaching, fluorescence recovery for each cell was calculated by dividing the ratio of fluorescence within the bleached region and the cell to the ratio of the fluorescence within these regions before bleach (i.e. the “steady state” distribution). The bleached region centroid was determined by the position of the maximum pixel value in the bleach frame, and the region radius was determined to be the full-width half maximum of the confocal spot intensity, ~3 pixels. The cell region was determined by first rotating the cell such that the long axis corresponded to the y-axis, then selected manually as a rectangular region. Due to incomplete photobleaching, the above normalized fluorescence recovery did not start at zero. To normalize each recovery to start at zero, we first estimated the zeroth timepoint by estimating the derivative at the first timepoint. Next, the minimum recovery value was subtracted from each timepoint and divided by the minimum recovery value subtracted from the theoretical maximum of one. Due to heterogeneity between cells, FRAP recoveries were averaged over all cells for each condition.

C. FRAP modeling

For FRAP simulations, differential equations describing changes in concentrations of the various states based on the kinetic rates described in the main text were solved numerically using the forward finite difference method for the bleached and fluorescent populations. Freely diffusing molecules were distributed equally between the simulated bleached region and the remainder of the cell. Residuals between the simulated FRAP recovery and the averaged FRAP recovery for each condition were minimized using the LSQNONLIN function in MATLAB version 2017a. Minimum values of 0 were imposed for all parameters. Initial guesses were randomized to ensure solutions were not due to convergence on a local minimum. For the EZRDM and AsiA overexpression conditions, two sets of parameters were found to be able to describe the data. For each condition, the parameter set that had the lowest squared residuals was chosen as the final model. Our model was reaction-dominated and ignored diffusion. To justify this, we plotted the fluorescence profile of each cell normalized against the prebleach fluorescence profile for multiple timepoints after photobleaching. For recoveries dictated by a reaction-dominant behavior, this normalized fluorescence profile should maintain its shape or change much more slowly than the rate of fluorescence recovery. For recoveries dictated by diffusion, the profile should smooth and flatten over time[56]. For all three conditions tested, fluorescence profiles maintained their shape over time, indicating the molecular behavior to be reaction-dominated, thus justifying the exclusion of diffusion from our model (Figure 2.2B).

D. SMT data collection

Cells were grown and harvested as previously described in Section IV.A. For the rpoC(I1309A) mutant, the MG1655 strain containing the pCH-rpoC(I1309A)-PAmCherry plasmid was induced with 1mM IPTG for 2 hours, then resuspended in media without IPTG and let grown at 30°C for one hours to allow for proper folding of the PAmCherry protein before harvest as described above. To view fluorescence, cells were illuminated with solid-state lasers at 405nm and 568nm wavelengths (Coherent). All SMT images were collected using 5ms exposure with 1.74ms of cycle time for a total frame length of 6.74ms. Three movies consisting of 2500 frames each were taken for each cell. To maximize data collection, the UV power was modulated between consecutive movies to maintain a similar activation rate across the imaging acquisition. No cell was imaged longer than five minutes to avoid phototoxicity effects.

E. SMT data analysis

For single molecule tracking analysis, tiff stacks of cell images were imported into the single molecule tracking software UTrack version 3.1[69] within the Matlab 2017a software. We performed the detection and tracking of molecules on individual movies using the Gaussian Mixture Model setting within the UTrack software. For detection, an α value of 0.01 was used; for frame-to-frame linking, 15 frames was the maximum time and 0 to 2 pixels (or up to 6 pixels with increased linking cost) was the maximum distance to link trajectories together. Linked trajectories were then filtered by their intensity (between 0.5-

1.5x the average single molecule intensity) and their localization within cells (i.e. any trajectory outside of cells was excluded) and exported as MATLAB files. Trajectories were analyzed using custom in-house code (available upon request) to assess their diffusive properties. Exact equations used to analyze the trajectories are explained in more detail in the supplemental information. To build the HMM, we installed and used the software vbSPT, version 1.1 on MATLAB version 2014a[50]. Cells were rotated such that the long cell-axis corresponded to the x-axis. We used only consecutive frame trajectories and only displacements in the x-direction for analysis.

F. Growth rate comparison between rpoC-PAmCherry and MG1655

To assess the growth and functionality of the rpoC-PAmCherry fusion strain, we compared the growth of the MG1655 parental strain to our rpoC-PAmCherry fusion strain at 24°C. Five cultures from distinct isolated colonies for each strain were inoculated in EZRDM media without antibiotics. Cells were shaken overnight at room temperature (~16 hours). Saturated cultures were diluted back 1:200 and the OD at 600nm was monitored over time. Three technical replicates of measurements for each timepoint were taken for each biological replicate. Data was pooled and averaged over all biological and technical replicates and the doubling time was determined by fitting to an exponential equation.

G. Western blot of RNAP

To assess whether the rpoC-PAmCherry protein was full-length and not proteolytically cleaved, we prepared cell lysates from an equivalent OD of cells for strains MG1655, MG1655::rpoC-PAmCherry, and MG1655//pCH-FtsZ-mCherry grown as previously described in Chapter 2, Section IV.B. Cell lysates were run on a 10% TBE protein gel (BioRad) and transferred at constant 25V for 2 hours at 4°C to a nitrocellulose membrane using a semi-dry transfer apparatus (GE). To ensure equal transfer, membranes were first blotted with Ponceau S stain to assess the transfer quality and rinsed with deionized water to remove the stain. Membrane blots were then incubated with 1x TBS + 0.1% Tween-20 solution (TBST) containing 5% w/v dry milk to block non-specific binding of the antibody for one hour at RT with gentle agitation, then washed three times for 5 minutes in TBST. Blocked membranes were then incubated with TBST + 1% w/v BSA (Sigma-Aldrich) with either a 1:50,000 dilution of anti-RpoC antibody (Biolegend clone NT73) or a 1:10,000 dilution of anti-mCherry antibody (Abcam ab167453) for one hour at RT with gentle agitation, then washed three times for 5 minutes in TBST. Membranes were then incubated in TBST with either 1:33,000 dilution of GAM-HRP secondary antibody (ThermoFisher 62-6520) for the anti-RpoC condition or GAR-HRP secondary antibody (BioRad #1705046) for the anti-mChery condition for one hour with gentle agitation, then washed five times for 5 minutes each in TBST. Membranes were then incubated with ECL solution diluted 1:5 in deionized water (BioRad #1705061) for 30 seconds, then wrapped in saran wrap and imaged using autoradiography film (Stellar Scientific BLI-810-100).

H. Co-immunoprecipitation of rpoC(I1309A)-PAmCherry with rpoB

Cell lysates were prepared as described in Secition IV.G for strains MG1655, MG1655//pCH-PAmCherry, and MG1655//pCH-rpoC(I1309A)-PAmCherry. Cell lysates were incubated with Protein G sepharose beads (Sigma-Aldrich P3296) with anti-RpoC antibody (Biolegend clone NT73) overnight at 4°C. Bead incubated lysates were then washed with wash buffer (20 mM Tris–HCl pH 7.9, 150 mM KCl, 1 mM $MgCl_2$) four times then incubated with 2x SDS sample buffer (1M Tris-HCl, pH 6.8, 50% glycerol, 10% SDS, 0.5% bromophenol blue, 0.5% β-mercaptoethanol) and boiled at 95% to elute protein. Samples at each stage were collected and run on a 10% TBE gel and transferred and detected as decribed in Section IV.G, but with detection using anti-RpoB antibody (BioLegend NT63) instead of anti-RpoC.

I. DNA staining and image collection

For DNA nucleoid imaging experiments, cells were incubated with Hoechst 33342 dye (bisbenzimide H33342 trihydrochloride, ThermoFisher H3570) at a concentration of 10 µg/mL for fifteen minutes at room temperature with cell agitation; cell culture tubes were wrapped in aluminum foil to prevent the dye from bleaching. Cells were then fixed in 3% paraformaldehyde in a PBS solution for 15 minutes at room temperature. Cells were then washed twice with PBS. Fixed and stained cells were mixed 1:1 with anti-fading solution (20% n-propyl gallate, 60% glycerol, 1x PBS) and adhered to #1.5 coverslips treated with 0.01% poly-lysine solution. Coverslips were sealed onto an imaging coverglass using clear nail polish and imaged on a GE OMX SR structured illumination microscope (excitation: 405, camera channel: 488, exposure 50ms, 5% of highest laser intensity). Images were reconstructed using the standard parameters for the GE OMX SR microscope within the GE SRx software.

V. Supplementary Notes

A. Determination of diffusion coefficients from single step displacements

To determine the diffusion coefficients and state occupation percentages present in the SMT data, we first calculated the single-frame displacement along the x-axis and plotted the histogram of their distribution binned into a number of bins equal to the square root of total number of displacements. This distribution resembles a Gaussian centered around zero; due to the nature of Brownian motion, displacements on average will be zero and the width of the distribution is related to the number of diffusion coefficients represented in the data and their values. We fit this distribution of single-frame displacements to Equation ((2.1).

$$p(r) = \sum_{i=1}^{n} \frac{p_i}{\sqrt{4\pi D_i t}} e^{-r^2/_{4D_i t}} \tag{2.1}$$

where n is the total number of diffusive states represented in the data, p_i is the ith state fraction, D_i is the ith state diffusion coefficient, and t is time.

Another method for determining the number, diffusion coefficients, and occupation percentages represented in SMT data is to calculate the single frame displacement in two dimensions for all trajectories, then construct the cumulative probability that the distribution will be less than or equal to a discrete value *x.* This cumulative probability distribution can be fit by using Equation ((2.2).

$$cdf(x) = 1 - \sum_{i=1}^{n} p_i e^{-x^2/_{4D_i t}} \tag{2.2}$$

where *n* is the total number of states represented in the data, *pi* is the *i*th state occupation fraction, *Di* is the *i*th state diffusion coefficient, and *t* is time.

B. Determination of spatial resolution

To determine the spatial resolution of our SMT experiments, we collected fluorescence images of fixed RNAP-PAmCherry cells using the same imaging collection procedure as described in Chapter 2, Section IV.D. Individual molecules were detected using the same procedure as the SMT data collection. For each detected molecule, the nearest neighbor distance in the consecutive frame up to 500nm was calculated. These nearest neighbor distances were binned and fit to an equation previously described[70].

Chapter 3: Probing the Molecular Mechanisms of HU-mediated Nucleoid Organization

I. Introduction

The *E. coli* genome must condense over a thousand fold to be packed into a cell of a few micrometers in size. This condensed genome, together with its associated proteins and RNAs, is termed the nucleoid and is organized into six macrodomains (MD) that are spatially and genetically isolated from each other[71-72]. The precise molecular mechanisms of how this nucleoid organization is achieved are not well understood, but many key players working at different length levels have been identified. A group of proteins known as nucleoid-associated proteins (NAPs), such as H-NS, IHF, Fis, and HU have been shown to play important roles[73]. HU is the most conserved NAP across eubacteria, and the second most abundant in *E. coli*, with around 30,000 copies per cell during middle exponential phase growth[74]. *E. coli* HU has two subunits, α and β, each at ~ 9 kDa, that form either HU$\alpha\alpha$ homodimers or HU$\alpha\beta$ heterodimers depending on the growth phase (HU$\beta\beta$ homodimers were negligible)[75]. It is well established that HU binds non-specifically to double-stranded (ds) chromosomal DNA, but can bind specifically with high affinity to distorted DNA such as kinked or cruciform DNA that are recombination intermediates[76].

HU has implications for numerous cellular processes including replication[77-78], transcription[79] and translation[80], many of which are mediated through its impact on the chromosomal structure. HU promotes chromosomal DNA contacts both on the short range (small DNA loops of a few Kbp)[81] and the long-range (~280kb[82]) through unknown mechanisms. While not lethal, deletion of both of HUα and HUβ results in significantly reduced compaction of the nucleoid[79], caused by the major loss of negative supercoiling that requires HU-constrained topological domains. Consequently, these cells are hypersensitive to gyrase inhibitor novobiocin[83]. While these functions of HU are critical, it is thought the main role of HU is in the general mediation of chromosome structure. A gain-of-function mutant of HU was found to drastically alter the transcriptional programming of the bacterial cell, wherein cells because spherical instead of their normal rod shape, had drastically condensed nucleoids, and expressed pathogenicity factors. Remarkably, this effect could be reversed by overexpressing WT HU exogenously, highlighting the ability of HU to modulate and reorganize the chromosome in response to various environmental cues and thus promoting differential transcriptional outputs[84]. However, the mechanism of how this organization is achieved has remained elusive.

HU binds several types of RNAs, including mRNAs, tRNAs and non-coding RNAs (ncRNAs) *in vivo*[80,85,86]. In particular, HU binds to a post-processed small ncRNA from a repetitive extragenic palindromic (REP) element, *REP325*[86-88]. Deletion of the *REP325* locus from the chromosome produced the same phenotype as the deletion of both HUα and β subunits, and exogenous introduction of one of the ncRNAs, named naRNA4 (nucleoid-associated RNA 4), was sufficient to rescue the phenotype. Additionally, chromosomal contacts between various REP elements in the E. coli genome were substantially reduced in cells that lacked either both subunits of HU or the *REP325* sequence[87]. Using a biochemical approach, it was found that HU:naRNA complexes are directly responsible for these particular chromosomal contacts, and that HU is not present in the final bridged product. This suggests HU acts as a catalytic chaperone, binding both cruciform DNA and the naRNA to bridge DNA-DNA contacts, an attractive model for the mediation of chromosome organization[88].

In addition to the RNA-binding chaperone activity of HU, it has been proposed that HU mediates chromatin organization through multimerization. In single molecule force pulling experiments, low concentrations of HU did not hinder DNA unzipping, but increased concentration caused an increase in DNA persistence length[89]. Similarly, crystal structures of HUαα and HUαβ dimers demonstrated that HU dimers were capable of interacting with each other, facilitating multiple HU dimer-dimers to bind linear DNA. Subsequently, the packing of these dimer-dimer interfaces promoted multimerization of HU, and predicted bundling of DNA via this multimerization. Intriguingly, the DNA networks possible by each of the two dimers differed substantially, with HUαβ able to form a more compact DNA network than HUαα. These observations were confirmed by SAXS, demonstrating the different binding activities of the two populations of HU dimers[90]. In this model, nucleoid condensation is achieved through the nonspecific binding activity of HU alone, not by RNA deposition. Given the copy number of HU (~30,000 in mid-log phase growth[75]), this nonspecific binding model of HU seems befitting. However, HU multimerization has not been directly examined in vivo. The naRNA chaperone activity of HU suggests its interactions with DNA may be transient. Conversely, the multimerization model suggests a stable network of HU:HU dimers and DNA, implying a long residence time on DNA.

For more insight into these possible molecular mechanisms, we utilize a fluorescent-protein labeled HU and monitor individual HU molecules to obtain their in vivo dynamics. We find that HU has two diffusive states, corresponding to DNA-bound and freely diffusing, respectively, and that the dynamics of state-switching is rapid, resulting in short dwell times of HU on the DNA. Using genetic mutations, we abolish the DNA-binding activities of HU and find a dramatic effect on HU dynamics and localization, effectively eliminating the ability of HU to be DNA-bound or nucleoid-associated. Abolishing structure-specific binding to gene regulatory loops had a less dramatic, but appreciable effect. Deleting the β subunit, we monitor HUαα homodimer dynamics and demonstrate differential dynamics to the wild-type mixed dimer condition, finding that HUαα likely mediates the non-specific binding of DNA while HUαβ seems to be more crucial for structure-specific binding. Finally, we find the RNA-binding activity of HU has modest effects on HU dynamics and likely plays a yet undefined role in chromosome or gene regulation homeostasis.

II. Results

To probe the dynamics of HUα and the molecular mechanism of HU-mediated nucleoid organization, we constructed several strains wherein PAmCherry was C-terminally fused to a WT or mutant version of the *hupA* gene and replaced at its endogenous locus. The strains used in this chapter are listed in Table 3.1 (below).

Table 3.1: strains used in this chapter, their genotype, and their respective reference.

Strain	Genotype	Reference
KB0026	BW25113::hupA-PAmCherry-cmR	[91]
ACL035	MG1655 galP::tetO$_6$ pDR-tetR- mCherry	This study
SCV148	MG1655::hupA-PAmCherry-cmR	This study
KB064	KB0026 ΔREP325	This study
SCV152	SCV ΔhupB	This study
SCV149	MG1655::hupA(K3A K18A K83A)-PAmCherry	This study
SCV153	SCV ΔhupB	This study
SCV146	MG1655::hupA(P63A)-PAmCherry	This study
SCV150	SCV146 ΔhupB	This study
SCV019	MG1655 ΔhupA	This study
KB061	MG1655 ΔhupB	Yale Genetic Stock
SCV020	MG1655 ΔhupA ΔhupB	This study

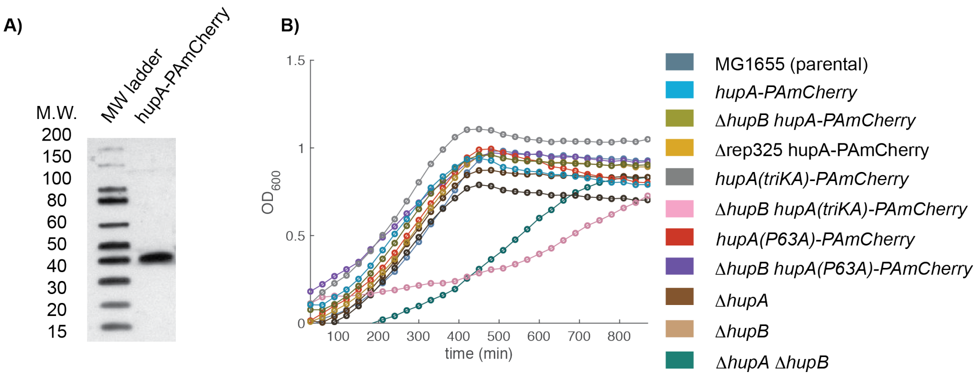
A. HU-PAmCherry exhibits two diffusive states

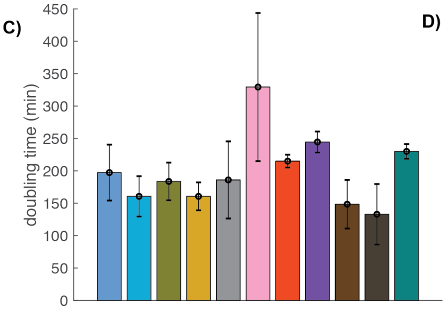
To probe the dynamics of HU in the nucleoid, we used an *E. coli* strain in which the chromosomal copy of *hupA* encoding for the α subunit of HU was replaced by a 3' tagged photoactivatible fluorescent protein fusion gene *hupA-PAmCherry*[91]. We chose to label the HUα instead of HUβ because HUα is incorporated into homodimeric HUα_2 and heterodimeric HU$\alpha\beta$ in log-phase

Table 3.2: growth rates of all strains imaged and the respective controls.

Each replicate represents cells from different isolate colonies; i.e. are true biological replicates. The times listed are the mean doubling time within log-phase growth.

Strain	Replicate 1	Replicate 2	Replicate 3
MG1655 (parental)	159 min	189 min	244 min
HUα-PAmCherry	149 min	137 min	196 min
ΔhupB HUα-PAmCherry	170 min	164 min	217 min
ΔREP325 hupA-PAmCherry	182 min	139 min	161 min
HUα(triKA)-PAmCherry	189 min	244 min	125 min
ΔhupB HUα(triKA)-PAmCherry	435 min	345 min	208 min
HUα(P63A)-PAmCherry	208 min	222 min	---
ΔhupB HUα(P63A)-PAmCherry	256 min	233 min	---
ΔhupA	122 min	175 min	---
ΔhupB	100 min	166 min	---
ΔhupA ΔhupB	238 min	222 min	---





D)

Mu phage assay

Strain	PFU of phage (Mu/mL)
MG1655 (parental)	4.6 x 10^{7}
HupA-PAmCherry	3.1 x 10^{7}
HupA-PAmCherry Δ*hupB*	4.2 x 10^{7}
SCV19 (Δ*hupB*)	3.7 x 10^{7}
SCV27 (Δ*hupA* Δ*hupB*)	<1.0 x 10^{3}*

* must be lower than experimentally tested

E)

mini-P1 plasmid replication assay

Strain	CFU of pUC19/μg DNA	CFU of mini-P1/μg DNA	Normalized mini-P1 CFU/μg DNA
MG1655 (parental)	1.2 x 10^{7}	3.8 x 10^{5}	3.8 x 10^{5}
HupA-PAmCherry	1.5 x 10^{8}	2.7 x 10^{6}	2.2 x 10^{5}
HupA-PAmCherry Δ*hupB*	4.7 x 10^{6}	1.3 x 10^{5}	3.3 x 10^{5}
SCV19 (Δ*hupB*)	9.5 x 10^{7}	7.9 x 10^{6}	1.0 x 10^{6}
SCV27 (Δ*hupA* Δ*hupB*)	1.4 x 10^{8}	< 2 x 10^{2} *	1.7 x 10^{1}*

* must be lower than experimentally tested

Figure 3.1: biochemical assays demonstrate the functionality of the HupA-PAmCherry fusion protein.

(A) Western blot of Huα-PAmCherry in cells grown in rich growth media. Lane 1: protein ladder. Lane 2: cell lysate blotted against PAmCherry with * marking the correct molecular size of 38 KD for the fusion protein. Image was cropped for size. (B) Growth curves of MG1655 (parental), *hupA-PAmCherry*, Δ*hupB hupA-PAmCherry*, *ΔREP325 hupA-PAmCherry*, *hupA(triKA)-PAmCherry*, Δ*hupB hupA(triKA)-PAmCherry*, *hupA(P63A)-PAmCherry*, *ΔhupB hupA(P63A)-PAmCherry*, Δ*hupA*, Δ*hupB*, and Δ*hupA* Δ*hupB* strains.. The doubling times are listed in Table 3.2. (C) Averaged doubling times, error bars represent standard deviation between biological replicates. (D) Function of HUα-PAmCherry fusion measured by the ability of Mu phage to infect each strain. PFU indicates the phage plaque formation unit. (E) Function of HUα-PAmCherry fusion measured by its ability to support mini-P1 plasmid replication. The colony formation unit (CFU) after transformation of the mini-P1 plasmid into each strain was normalized against each strain's relative competency (CFU of a PUC19 plasmid transformation) compared to that of the wild type MG1655 parental strain.

growth (HUβ_2 homodimers are not detectable in log-phase *E. coli*)[74]. We verified that the fusion protein HUα-PAmCherry was expressed at the expected full length as the sole cellular source of HUα and supported wild-type (WT)-like growth (Figure 3.1ABC). Because a single deletion of *hupA* or *hupB* is able to function similarly as WT cells, but a double deletion has detrimental effects, we constructed the same strain but with *hupB* deleted. In this strain, we observed HU-dependent mini-P1 plasmid replication and Mu phage growth in the fusion protein strain nearly indistinguishable from that of the parental strain[86,87] (Figure 3.1DE). These results suggested HUα-PAmCherry could replace the endogenous HUα for its function.

To perform single molecule tracking on HUα-PAmCherry, we imaged mid log-phase live *E. coli* cells grown in EZ rich defined media on agarose gel pads. We activated individual HUα-PAmCherry molecules using a low level of 405 nm UV light and monitored their cellular positions with a frame rate of ~150 Hz (Δt = 6.74ms) using laser excitation at 568 nm. This imaging rate allowed us to determine confidently the diffusion coefficients of individual molecules up to 3 $\mu m^2/s$. In Figure 3.2A we showed a few representative trajectories superimposed with the corresponding bright-field image of the cell.

Individual trajectories displayed different diffusive properties: some were confined within small regions (Figure 3.2A, blue trajectories) and some traversed larger cell areas (Figure 3.2A, red trajectories). Plotting the apparent diffusion coefficients of individual trajectories indeed suggested that there existed a

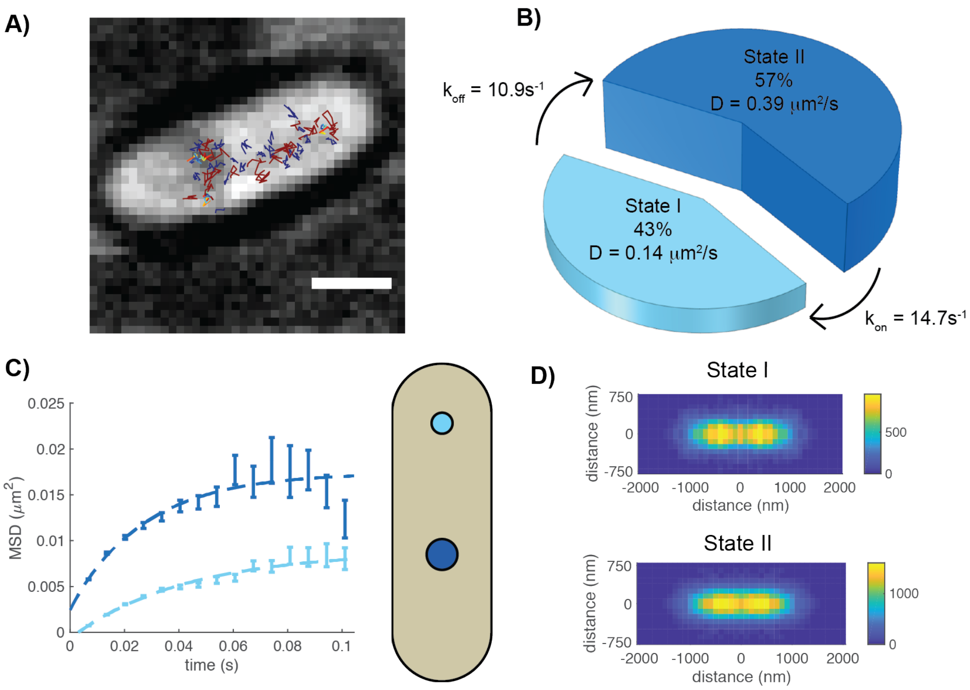


Figure 3.2: Single molecule tracking (SMT) of HUα-PAmCherry in live *E. coli* cells grown in rich media showed two diffusive states, rapid kinetics, and a nucleoid localization.

(A) Representative SMT trajectories superimposed on top of the corresponding brightfield image of the cell (gray) with scale bar = 1 micron. Blue trajectories represent State I and red represent State II. Molecules that transitioned between the two states are colored in rainbow. (B) Two diffusive states of HUα-PAmCherry molecules with respective population percentage, transition rates, and diffusion coefficients as identified by the HMM. (C) The mean squared displacement (MSD) of State I (light blue) and State II (dark blue) HUα-PAmCherry trajectories as a function of time; the plateau signifies confined diffusion with State I having confinement zone of 230 nm and State I with 300 nm. These values are represented as bubbles within a typical cell to the right of the graph. (D) Two-dimensional (2D) histograms of State I (top) and State II (bottom) HUα-PAmCherry localizations in a standard 4 μm x 1.5 μm cell. Because of the symmetry of the cell shape in both long and short axes, we calculated the absolute displacement of each HUα-PAmCherry localization to the center of the cell, normalized its long axis displacement to the standard cell length, and duplicated the quartile cell histogram along both the long and short axes to produce a full-sized 2D histogram of HUα-PAmCherry distribution. The bin size of the 2D histogram was 100 x 100 nm. The color bar indicated localization numbers used in each bin. A total number of 23,236 localizations for State I and 37,196 for State II were used for the localization distributions.

heterogeneous distribution of HUα-PAmCherry molecules with a wide range of diffusion coefficients (Figure 3.3A). To analyze the different populations quantitatively, we used a Bayesian-based Hidden Markov Model (HMM)[50], which

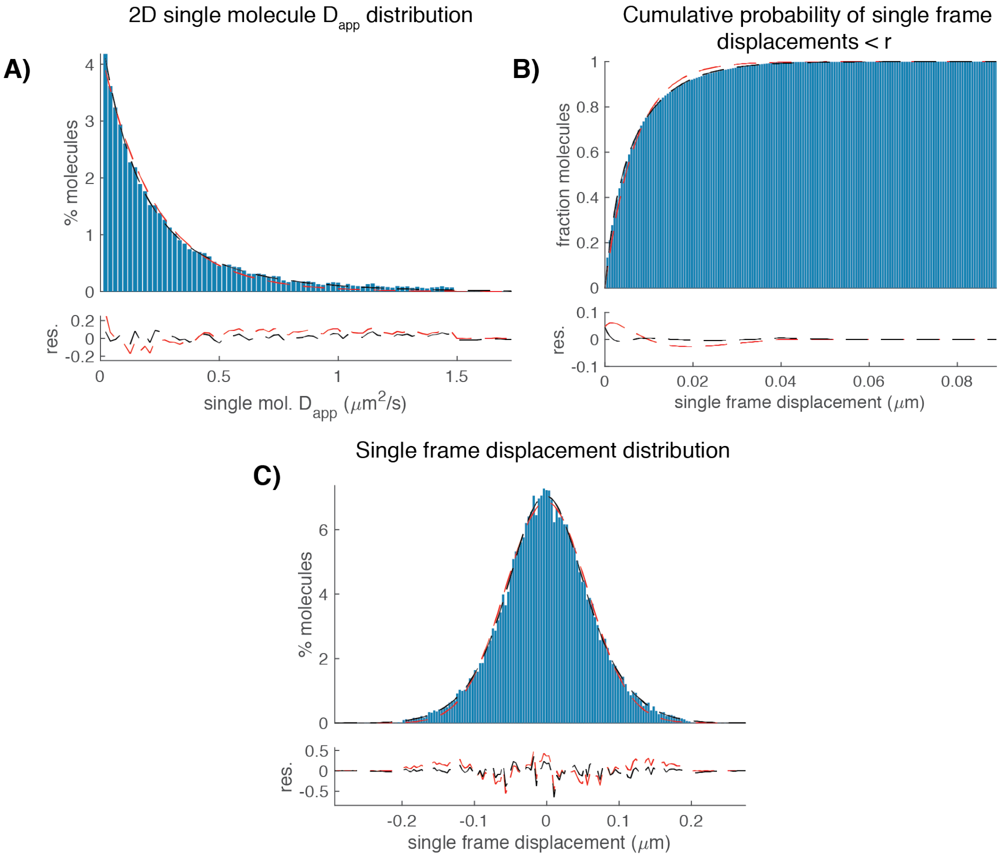


Figure 3.3: different methods to determine the number, diffusion coefficients, and occupation percentages of states for WT HUα-PAmCherry.

(A) For each individual displacement, the diffusion coefficient was calculated in two dimensions and binned. (B) Single frame displacements in two dimensions were calculated and the cumulative probability of displacements less than or equal to a value *r* was determined. (C) Single frame displacements were calculated along the long x-axis of the cell and binned. Red lines indicate single population fits, while black lines indicate two population fits. Residuals for each fit are plotted below each distribution. Equations used to fit each distribution are listed in Section V. Fit parameters determined from the fits are listed in Table 3.3.

robustly determines the optimal number of diffusive states represented in the data, the diffusion coefficients of those states, and the occupation percentage of molecules in each state. We found that the data fit best to a two-state model (Figure 3.2B), with diffusion coefficients of State I $D_1 = 0.14 \pm 0.004$ $\mu m^2/s$, and State II $D_2 = 0.39 \pm 0.006$ $\mu m^2/s$, ($\mu \pm$ s.e.m., n = 60,432 displacements). The occupation percentages were ~45% and ~ 55% for State I and II respectively.

Table 3.3: fit parameters from the various classical methods of determining the diffusive states represented in SPT data for WT HUα-PAmCherry.

The methodologies are 1D SFD, i.e. one dimensional single frame displacement distribution from Figure C;, the CDF of the cumulative displacement probability function from Figure B, and 2D ADD, i.e. the two-dimensional apparent diffusion coefficient distribution from Figure A. Values listed are means ± 95% confidence intervals.

HUα-PAmCherry cells under rich growth media conditions				
Methodology	D_1 (μm^2/s)	p_1 (%)	D_2 (μm^2/s)	p_2 (%)
1D SFD	0.14 ± 0.005	43 ± 5	0.38 ± 0.01	57 ± 6
CDF	0.12 ± 0.006	36 ± 3	0.35 ± 0.007	64 ± 3
2D ADD	0.13 ± 0.07	26 ± 4	0.29 ± 0.08	74 ± 4

Using several other classic methods[48], we obtained the same two states with essentially identical diffusion coefficients and population percentages (Figure 3.3 and Table 3.3). Based on these results, we concluded that there were at least two HUα-PAmCherry populations in the nucleoid of live *E. coli* cells that could be distinguished based on their different diffusive behaviors and cellular distributions.

B. The two diffusive states reflect HU-PAmCherry's interactions with chromosomal DNA

We reasoned that the two diffusive states of HUα-PAmCherry likely resulted from its interactions with chromosomal DNA in the nucleoid[76,80,92]. The slow diffusive State I likely represented HUα-PAmCherry molecules bound to chromosomal DNA, whereas fast diffusive State II HUα-PAmCherry molecules the unbound population diffusing in the nucleoid. To support our theory that State I represented DNA-bound HU, we tracked the diffusion of a chromosomal DNA segment labeled with six *tetO* operator sites (*tetO*^{6}) tightly bound by TetR-

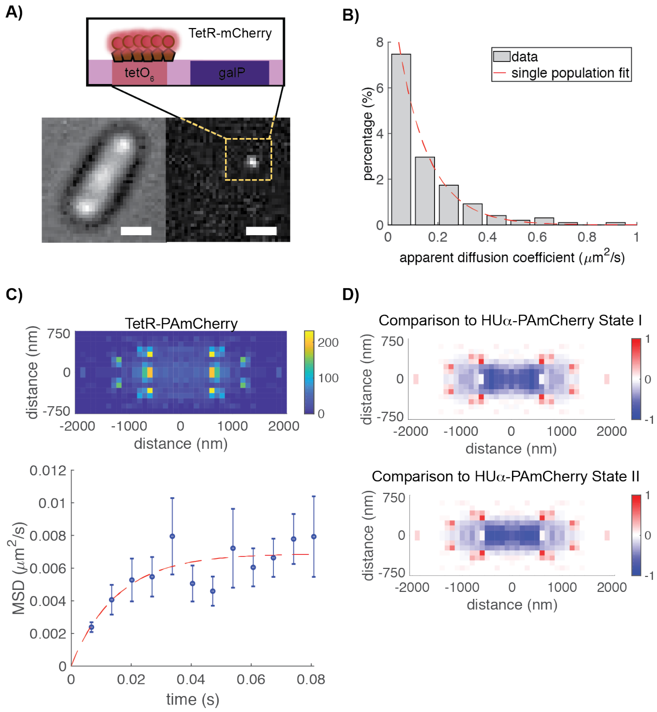


Figure 3.4: single molecule tracking of a DNA locus exhibited similar diffusive behavior as State I HUα-PAmCherry.

(A) Schematic of FROS with representative fluorescent and brightfield images of a cell. Scale bar, 1 μm, (B) Distribution of apparent diffusion coefficients of individual TetR-mCherry molecules. (C) 2D histogram of all TetR-PAmCherry localizations within a normalized cell. Coordinates are binned into 100x100nm bins. (D) Log2 comparison of normalized localizations between the DNA locus and State I HUα-PAmCherry. (E) MSD of DNA loci displays confinement with similar diffusive domain value to State I HUα-PAmCherry. Values are μ ± s.e.m.

mCherry fusion protein molecules using the same cell growth and imaging conditions as HUα-PAmCherry (Figure 3.4A).

We observed that TetR-PAmCherry formed one or two distinct spots inside cells and diffused with an average diffusion coefficient of 0.09 ± 0.04 $\mu m^2/s$ (μ ± s.e.m, n = 1808 displacements, Figure 3.4B), similar to that of State I HU-PAmCherry molecules and what was previously reported on labeled *E. coli* chromosomal foci (~ 0.1 $\mu m^2/s$, [93]). Note that compared to other fluorescent repressor-operator systems (FROS) using tandem arrays of hundreds of DNA binding sites[94], the small foot print of the *tetO*^{6} site (~ 200 bp) enabled us to pinpoint the position of the labeled chromosomal DNA with high accuracy and negligible perturbations [95]. The reconstructed cellular localizations of TetR-mCherry showed a much more narrow distribution compared to that of State I HUα-PAmCherry (Figure 3.4CD), consistent with the expected specific, high-affinity binding of TetR to the only *tetO*^{6} site(s) in the cell, and the nonspecific binding of HU to chromosomal DNA. Both HUα-PAmCherry State I and State II molecules have MSD values that deviate from the linear behavior of normal Brownian diffusion, instead plateauing at longer time lags, indicative of confined diffusion. Fitting the MSD to a modified equation that assumes a finite boundary, we found the confinement size of State I to be ~ 230 nm in diameter. Unexpectedly, State II also experienced confinement at ~ 300nm, and both states showed similar localization distributions within a normalized cell (Figure 3.2D), suggesting that if HU unbinds DNA, it likely remains close to its prior binding site. For State II molecules, while the diffusion coefficient of ~ 0.4 $\mu m^2/s$ was three-fold larger than that of State I molecules, it was about one order of magnitude smaller than what would be expected from a freely diffusing, non-DNA-interacting molecule of comparable size in the *E. coli* cytoplasm[31]. Despite this observation, this diffusive behavior is consistent with nucleoid-localized diffusing populations of other DNA-binding proteins in *E. coli*, such as lacI[41], RNA polymerase[60], and UvrA[64]. This state differs from that of truly freely diffusing proteins outside of the dense nucleoid; RNA polymerase, for example, was found to diffuse at ~ 2 $\mu m^2/s$ at DNA-free end-caps of replication-stalled cells[60]. This suggests that State II HU may be in some way loosely associated with DNA as opposed to truly freely diffusing.

C. HU interacts with chromosomal DNA with transient kinetics

One powerful aspect of SMT is the ability to obtain transition probabilities of molecules between different diffusive states. These transition probabilities provide crucial information regarding the kinetics of state-switching, such as the binding and unbinding rates of a protein molecule to its target site, and the lifetime of a particular functional state of the molecule. In some trajectories, we indeed observed events that likely represented the switching between the slow and fast diffusive states of individual HUα-PAmCherry molecules, suggesting the binding and unbinding events of HU on chromosomal DNA (Figure 3.2A, rainbow trajectory). Using the HMM, we found that HUα-PAmCherry molecules switched from the slow diffusing State I to the fast diffusing State II at an apparent rate of k_{12} (or k_{off}) = 10.9 s^{-1}, and from State II to State I at k_{21} (or k_{on}) = 14.9 s^{-1} (Figure 3.2B). The dwell time of State I and II were found at 100 ± 15 ms and 74 ± 7 ms, respectively. These dwell times were significantly shorter than what would be expected for the specific binding of most DNA binding proteins[96], or the known nanomolar affinity of HU to nicked, cruciform, and kinked DNA[76,85]. These results corroborated the relatively wide cellular distributions of State I HUα-PAmCherry molecules compared to that of TetR-mCherry (Figure 3.4D), suggesting that

HUα-PAmCherry interacts with chromosomal DNAs transiently and nonspecifically across the nucleoid.

D. HU dynamics depend on non-specific DNA-binding

The transitory interactions of HU with the DNA are consistent with prior biochemical data signifying its low affinity to linear dsDNA. Crystal structures of HU demonstrated that positively-charged lysine residues along the outer surface of HU dimers mediate its interactions to negatively-charged linear dsDNA. Thus, mutation of these lysine residues to the neutral amino acid alanine should eliminate non-specific binding of HU to DNA. To investigate HU dynamics when non-specific binding was abolished, we constructed a strain in which three lysines on the surface of the HUα subunit were mutated to alanines (K3A K18A K83A). We C-terminally fused *PAmCherry* to the mutant *hupA* gene at its endogenous locus, hereto referred to as HUα(triKA)-PAmCherry. We tracked the dynamics of HUα(triKA)-PAmCherry using the same growth and imaging conditions as the WT HUα-PAmCherry cells.

HUα(triKA)-PAmCherry molecules displayed dramatically altered dynamics as compared to the WT condition. We found that there were two diffusive states, with diffusion coefficients of 0.17 $\mu m^2/s$ and 1.08 $\mu m^2/s$ (n = 193,678 displacements from 61 cells) (Figure 3.5). We again corroborated the results from the HMM with classical methods (Figure 3.6). State I HUα(triKA)-PAmCherry diffused similarly to State I WT HUα-PAmCherry, suggesting the function of this diffusive state is similar, i.e. DNA-bound. However, only 5% of

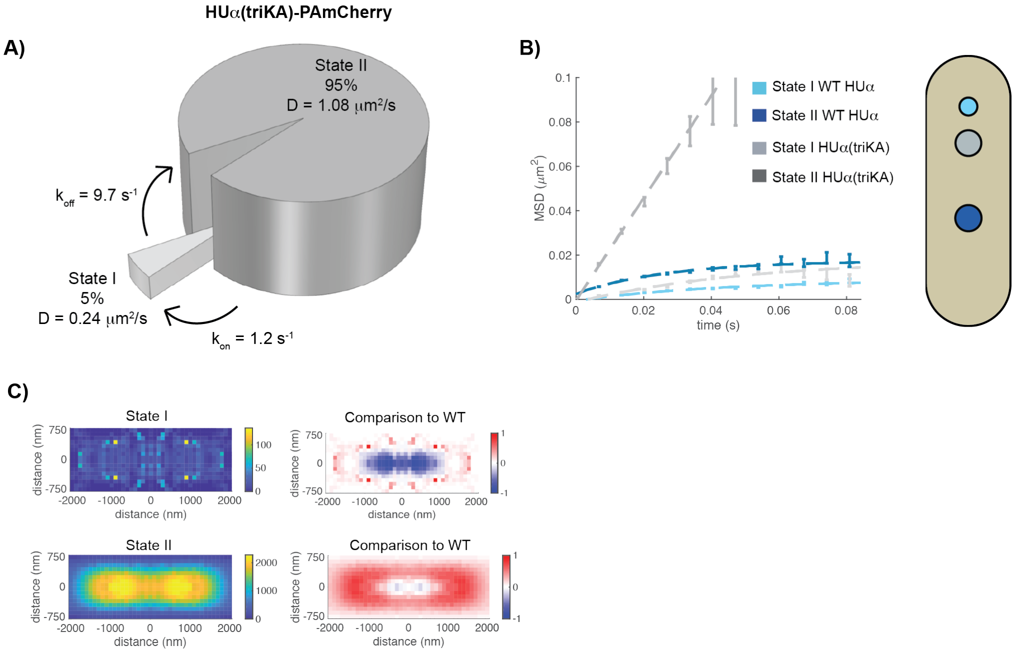


Figure 3.5: disabling HUα-PAmCherry from binding non-specific DNA through a triple lysine mutation exhibited dramatically altered dynamics and localization.

A) Two diffusive states of HUα(triKA)-PAmCherry with respective diffusion coefficients, transition rates, and population percentages as identified by the HMM. (B) MSD as a function of time for State II molecules (light and dark gray) compared to State I and II molecules in the WT condition (light and dark blue). Representative confinement sizes relative to a typical cell are shown to the right. (C) 2D histograms of State I and II HUα(triKA)-PAmCherry localizations within a normalized 4 μm x 1.5 μm cell. The bin size of the histogram was 100 x 100 nm. A total of 2,256 and 191,422 localizations for State I and II respectively were used for histogram construction. Comparison to the WT condition is shown in the right-hand column. Values represent the log2-fold change in normalized density.

HUα(triKA)-PAmCherry molecules were in State I, close to the error in detecting diffusive states using our HMM software[50], suggesting this mutant HU lost most to all of its ability to bind to nsDNA.

Previously we hypothesized State II represented HUα-PAmCherry molecules diffusing within the nucleoid while loosely associating DNA. The increase in diffusion coefficient of State II for HUα(triKA)-PAmCherry suggests these molecules undergo less “sticky” diffusion and are more freely diffusing. Plotting the 2D histogram of HUα(triKA)-PAmCherry shows that while a bi-lobed

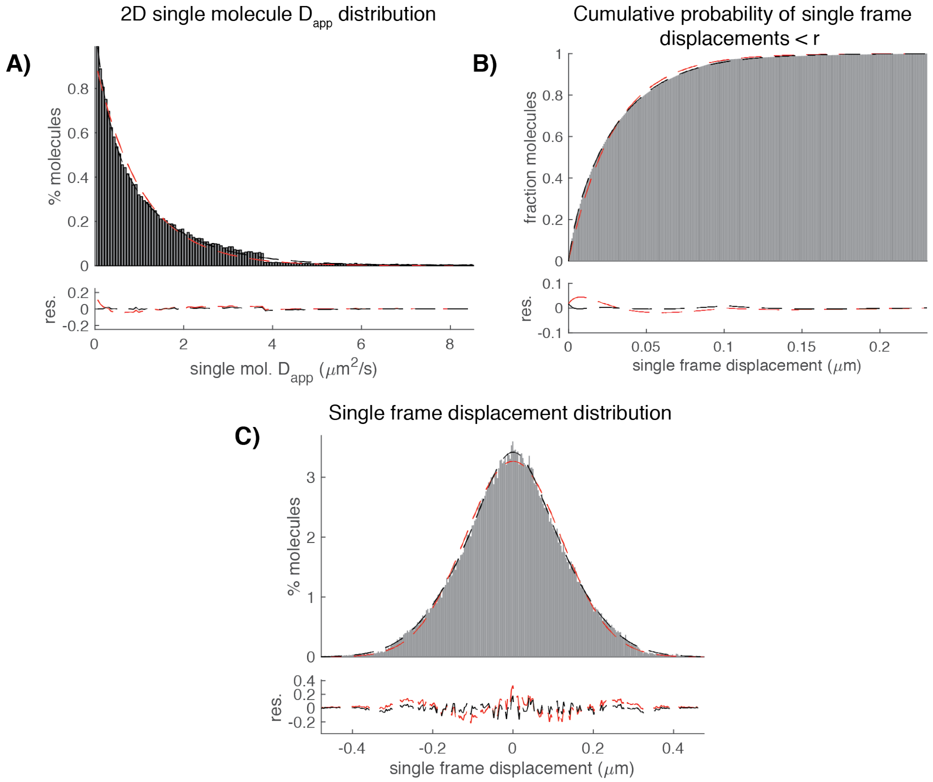


Figure 3.6: different methods to determine the number, diffusion coefficients, and occupation percentages of states for HUα(triKA)-PAmCherry.

(A) For each individual displacement, the diffusion coefficient was calculated in two dimensions and binned. (B) Single frame displacements in two dimensions were calculated and the cumulative probability of displacements less than or equal to a value *r* was determined. (C) Single frame displacements were calculated along the long x-axis of the cell and binned. Red lines indicate single population fits, while black lines indicate two population fits. Residuals for each fit are plotted below each distribution. Equations used to fit each distribution are listed in Section V. Fit parameters determined from the fits are listed in Table 3.4.

distribution can be discerned, the distribution is much widened compared to the WT condition (Figure 3.5C, comparison to WT is on the right hand column). Additionally, HUα(triKA)-PAmCherry molecules had a dramatic increase in confinement zone size, from 300nm in diameter to well over a micron in diameter for State II molecules (Figure 3.5B, compare dark blue with dark gray), consistent

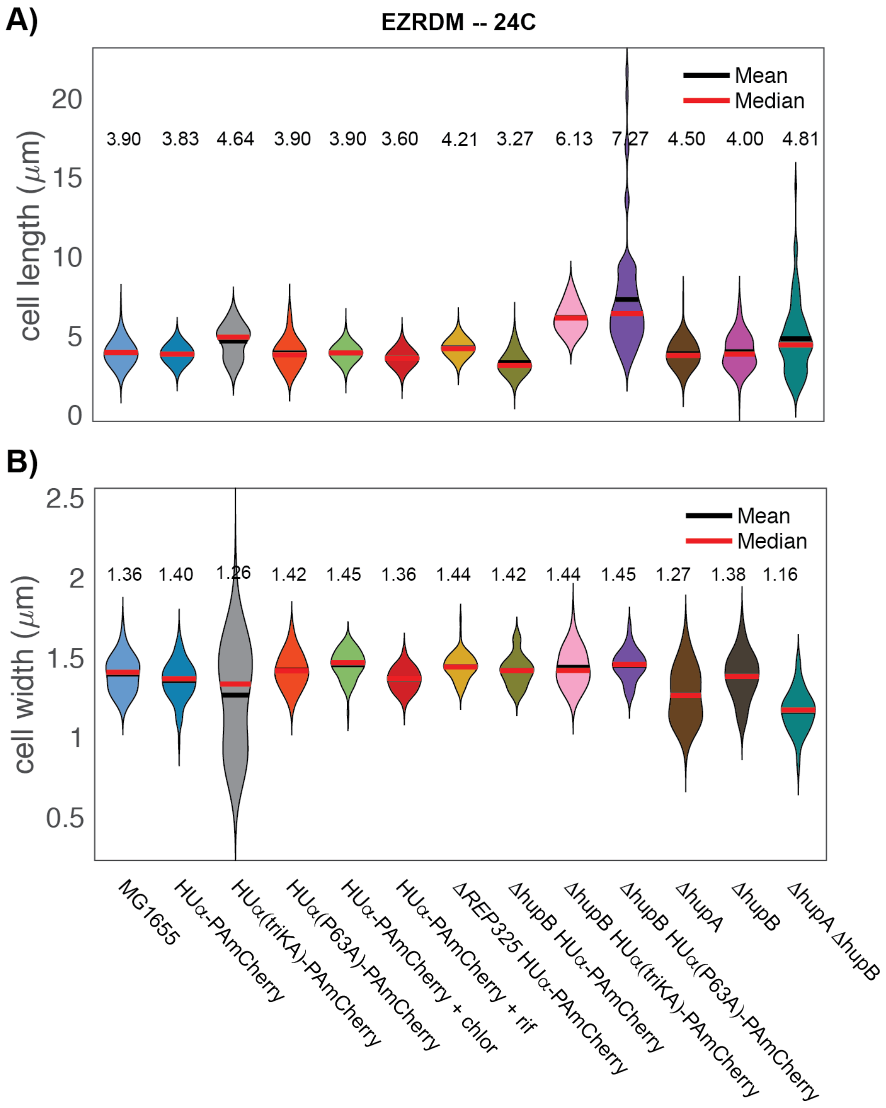


Figure 3.7: cell length and width of all imaging strains and controls at 24°C.

(A) Cell lengths and (B) widths of MG1655, MG1655::hupA-PAmCherry, MG1655::hupA(triKA)-PAmCherry, MG1655::hupA(P63A)-PAmCherry, MG1655::hupA-PAmCherry treated with chloramphenicol, MG1655::hupA-PAmCherry treated with rifampicin, ΔREP325 hupA-PAmCherry, ΔhupB hupA-PAmCherry, ΔhupB hupA(triKA)-PAmCherry, ΔhupB hupA(P63A)-PAmCherry, ΔhupA, ΔhupB, and ΔhupA ΔhupB strains. Mean values are represented by black bars with values listed above each violin bar, median values are represented by red bars. Thickness of the violin bar indicates the relative percentage of cells with that respective cell length or width.

with the idea HUα(triKA)-PAmCherry molecules are unable to “see” the nucleoid due to the loss of electrostatic interactions with DNA. SIM imaging of the nucleoids in this strain suggest no significant deviations from WT in terms of

Table 3.4: fit parameters from the various classical methods of determining the diffusive states represented in SPT data for HUα(triKA)-PAmCherry.

The methodologies are 1D SFD, i.e. one dimensional single frame displacement distribution from Figure C;, the CDF of the cumulative displacement probability function from Figure B, and 2D ADD, i.e. the two-dimensional apparent diffusion coefficient distribution from Figure A. Values listed are means ± 95% confidence intervals.

HUα(triKA)-PAmCherry cells				
Methodology	D_1 (μm^2/s)	p_1 (%)	D_2 (μm^2/s)	p_2 (%)
1D SFD	0.25 ± 0.08	11 ± 2	1.32 ± 0.08	89 ±17
CDF	0.18 ± 0.03	12 ± 1	1.21 ± 0.03	88 ± 1
2D ADD	0.28 ± 0.08	12 ± 0.1	1.30 ± 0.1	88 ± 0.1

volume, perhaps due to HUβ compensating for the defects in HUα. The range in cell lengths for the triple lysine mutant seem to be bimodal, with a population resembling wild-type length and another that are elongated (Figure 3.7A). Additionally, cell widths were overall short but displayed a wide range of values (Figure 3.7B). Together, this suggests that dosage of the triple lysine HUα subunit affects cell phenotype, implying the triple lysine mutant has a dominant negative effect.

While the role of HU is presumed to be primarily binding non-specifically to DNA, HU can bind structure-specifically, such as at the *gal* operon during repressor loop formation[97,81,98]. A nearly universally conserved proline P63 located in the beta-branch arm of each HU subunit intercalates the minor groove of DNA, stabilizing loop and bent conformations[99,90]. Thus, mutation of P63 to alanine should abolish site-specific HU binding while keeping non-specific binding of HU intact, as the nonspecific interactions do not require P63. To test

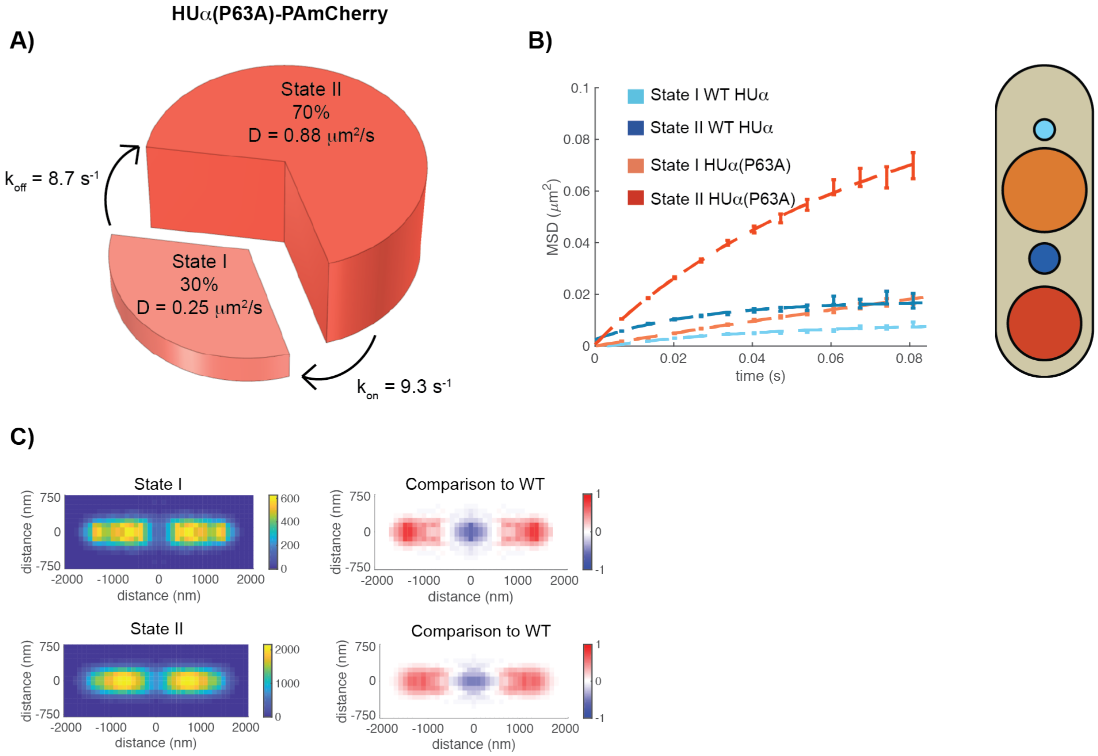


Figure 3.8: mutating a conserved proline residue responsible for site specific binding moderately altered HU dynamics.

(A) Two diffusive states of HUα(P63A)-PAmCherry with respective diffusion coefficients, transition rates, and population percentages as identified by the HMM. (B) MSD values for State I and II HUα(P63A)-PAmCherry molcules (light and dark orange) compared to the WT condition (light and dark blue). Representative confinement sizes relative to a typical cell are shown to the right. (C) 2D histograms of State I and II HUα(P63A)-PAmCherry localizations within a normalized 4 μm x 1.5 μm cell. The bin size was 100 x 100 nm. A total of 21,029 and 69,377 localizations for State I and State II respectively were for histogram construction. Comparison to the WT condition is shown in the right-hand column. Values represent the log2-fold change in normalized density.

whether this structure-specific DNA-binding ability had an influence on HU dynamics and localization, we again C-terminally fused *PAmCherry* to the *hupA(P63A)* mutant gene.

Using the HMM analysis, we found that HUα(P63A)-PAmCherry molecules had two diffusive states with diffusion coefficients of 0.25 ± μm2/s and 0.88 ± μm2/s with occupation percentages of 32 ± % and 68 ± % for State I and II respectively (n = 90,406 displacements from 60 cells) (Figure 3.8D). These

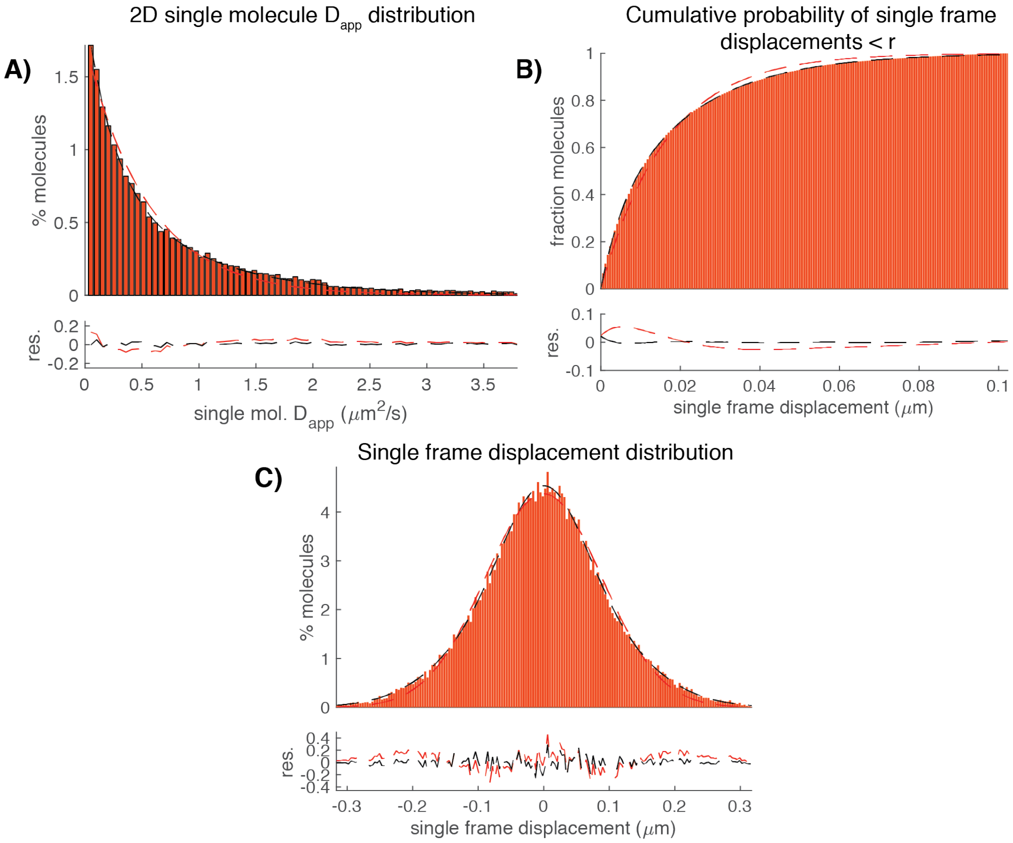


Figure 3.9: different methods to determine the number, diffusion coefficients, and occupation percentages of states for HUα(P63A)-PAmCherry.

A) For each individual displacement, the diffusion coefficient was calculated in two dimensions and binned. (B) Single frame displacements in two dimensions were calculated and the cumulative probability of displacements less than or equal to a value *r* was determined. (C) Single frame displacements were calculated along the long x-axis of the cell and binned. Red lines indicate single population fits, while black lines indicate two population fits. Residuals for each fit are plotted below each distribution. Equations used to fit each distribution are listed in Section V. Fit parameters determined from the fits are listed in Table 3.5.

diffusive states determined by the HMM were verified using the three classical methods (Figure 3.9). The increase in diffusion of State I suggests some loss of the ability of HU to bind stably to DNA. Additionally, the increase in diffusion of State II implies a lowered affinity of unbound HU molecules for loose association to DNA. HU dynamics remained transitory with State I having dwell time of 129 ms and State II with 114 ms, with transition rates k_{off} and k_{on} fairly similar to WT

Table 3.5: fit parameters from the various classical methods of determining the diffusive states represented in SPT data for HUα(P63A)-PAmCherry.

The methodologies are 1D SFD, i.e. one dimensional single frame displacement distribution from Figure C;, the CDF of the cumulative displacement probability function from Figure B, and 2D ADD, i.e. the two-dimensional apparent diffusion coefficient distribution from Figure A. Values listed are means ± 95% confidence intervals.

HUα(P63A)-PAmCherry cells				
Methodology	D_1 ($\mu m^2/s$)	p_1 (%)	D_2 ($\mu m^2/s$)	p_2 (%)
1D SFD	0.26 ± 0.08	28 ± 4	0.89 ± 0.17	72 ± 10
CDF	0.19 ± 0.02	26 ± 2	0.79 ± 0.02	74 ± 4
2D ADD	0.24 ± 0.06	26 ± 1	0.83 ± 0.1	74 ± 1

cells (Figure 3.8D). Both States I and II had increased confinement sizes compared to WT, with State I having confinement size of 876 ± 6 nm and State II with 796 ± 1 nm in diameter (Figure 3.8E, orange lines). Despite this increase in confinement zone size, States I and II had cellular distributions similar to WT cells, but with a relative depletion of localizations at mid-cell and enhancement near the polar edge of the nucleoid (Figure 3.8F). The increase in diffusion coefficient for both states might suggest a partial loss of function, yet the cells exhibited normal growth rates and cell length distributions (Figure 3.1B and Figure 3.7). Additionally, the localizations of HUα(P63A)-PAmCherry were more narrowly distributed (Figure 3.8F), unlike HUα(triKA)-PAmCherry, suggesting this P63A mutant subunit has at least partial ability to bind non-specifically to DNA.

E. HU dynamics are affected by loss of RNAs but not necessarily nucleoid reorganization

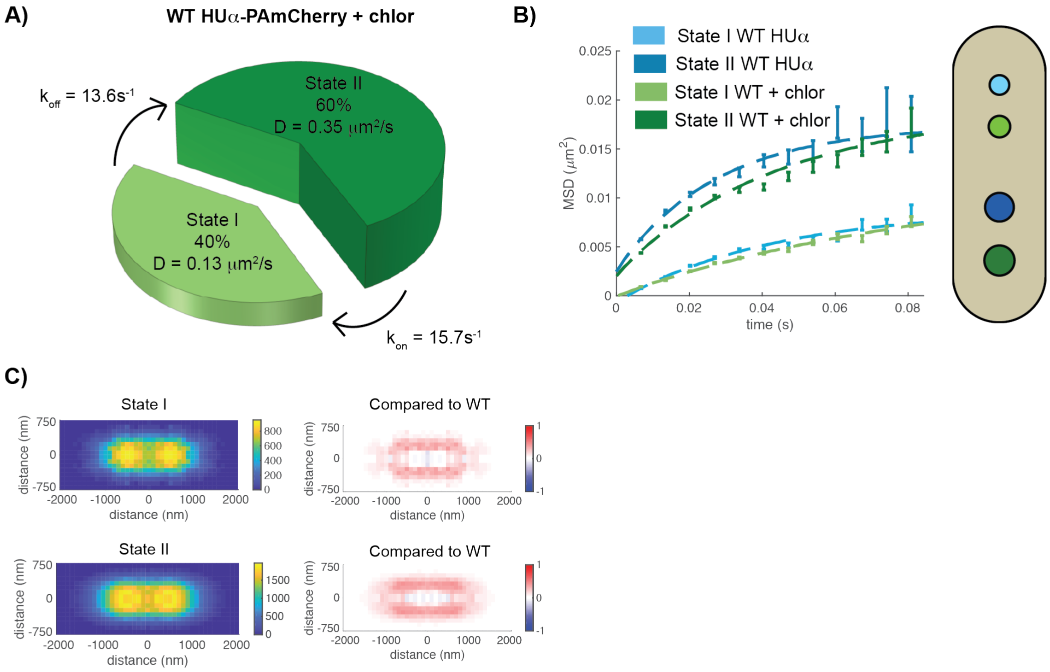


Figure 3.10: compaction of the nucleoid has minimal effect on HUα-PAmCherry dynamics.

(A) Two diffusive states of HUα-PAmCherry when treated with chloramphenicol with respective diffusion coefficients, transition rates, and population percentages identified from the HMM. (B) MSD values for chloramphenicol-treated cells (light and dark green) compared to the untreated condition (light and dark blue). Representative confinement sizes for each state and condition relative to a typical cell are shown to the right. (C) 2D histogram of State I and II HUα-PAmCherry localizations under chloramphenicol treatment within a normalized 4 µm x 1.5 µm cell. The bin size was 100 x 100 nm. A total of 33,610 and 75,236 localizations for States I and II respectively were used for histogram construction. Comparison to the untreated condition is shown in the right-hand column. Values represent log2-fold change in normalized density.

The altered dynamics of the DNA-binding mutants of HU may be an indirect effect of altered chromatin organization resulting from a changed transcriptional program, as previously demonstrated[79,84,100]. To probe the possibility that chromosome organization may influence HU dynamics, we altered the nucleoid size through two previously characterized drug perturbations, chloramphenicol and rifampicin, which condense and expand the nucleoid

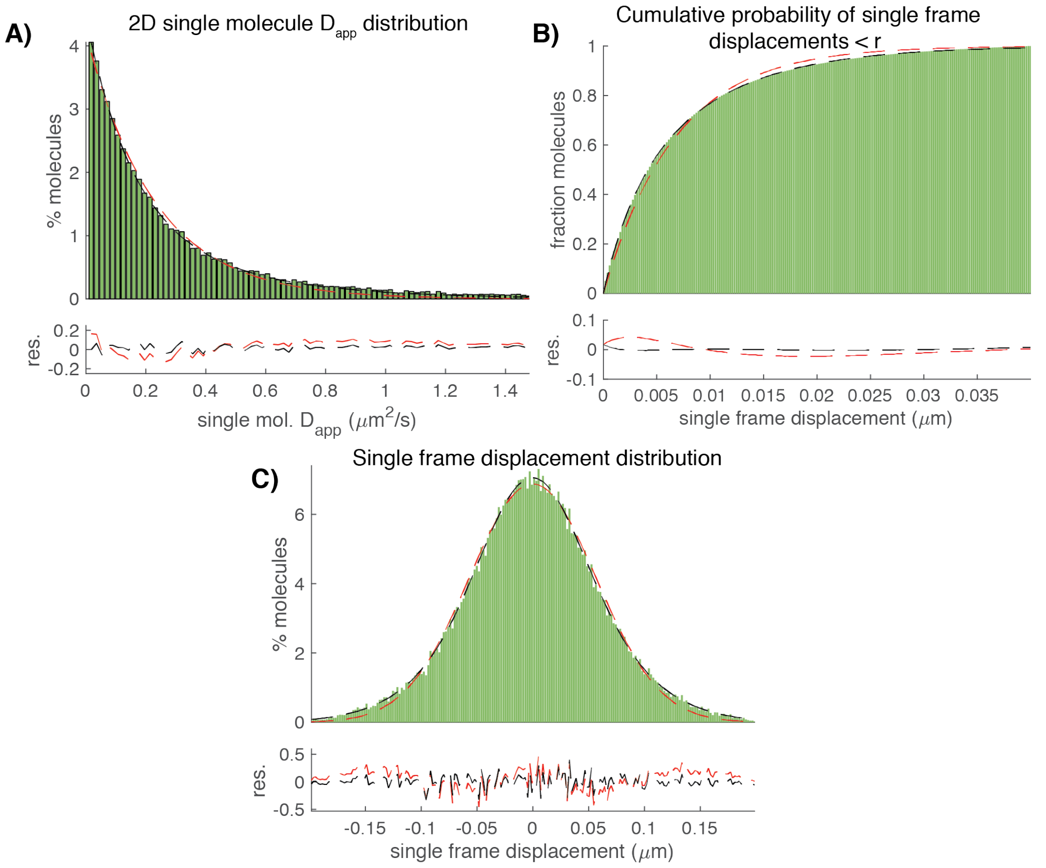


Figure 3.11: different methods to determine the number, diffusion coefficients, and occupation percentages of states for HUα-PAmCherry treated with chloramphenicol.

A) For each individual displacement, the diffusion coefficient was calculated in two dimensions and binned. (B) Single frame displacements in two dimensions were calculated and the cumulative probability of displacements less than or equal to a value *r* was determined. (C) Single frame displacements were calculated along the long x-axis of the cell and binned. Red lines indicate single population fits, while black lines indicate two population fits. Residuals for each fit are plotted below each distribution. Equations used to fit each distribution are listed in Section 5. Fit parameters determined from the fits are listed in Table 3.6.

respectively[59]. Chloramphenicol is a global translation inhibitor[101] that condenses the nucleoid due to the halt of transertion of membrane proteins into the inner membrane, a major nucleoid expansion force[59]. At this reduced nucleoid volume, we observed diffusive states and occupation percentages that were essentially the same as compared to untreated cells (Figure 3.10A) (n = 108,846

Table 3.6: fit parameters from the various classical methods of determining the diffusive states represented in SPT data for HUα-PAmCherry cells treated with chloramphenicol.

The methodologies are 1D SFD, i.e. one dimensional single frame displacement distribution from Figure C;, the CDF of the cumulative displacement probability function from Figure B, and 2D ADD, i.e. the two-dimensional apparent diffusion coefficient distribution from Figure A. Values listed are means ± 95% confidence intervals.

HUα-PAmCherry cells treated with chloramphenicol				
Methodology	D_1 ($\mu m^2/s$)	p_1 (%)	D_2 ($\mu m^2/s$)	p_2 (%)
1D SFD	0.14 ± 0.002	43 ± 2	0.4 ± 0.006	57 ± 3
CDF	0.11 ± 0.007	30 ± 3	0.32 ± 0.006	70 ± 3
2D ADD	0.14 ± 0.04	37 ± 3	0.34 ± 0.08	63 ± 3

displacements from 61 cells). Although the k_{off} value slightly increased, the k_{on} value remained nearly identical (Figure 3.10A). Moreover, the confinement of State I and State II molecules was again virtually the same as untreated cells (Figure 3.10B, Table S5). Consequently, HU localizations mimicked untreated cells, with some expansion of localizations along the y-axis for chloramphenicol-treated cells (Figure 3.10C). Ultimately, this data suggested nucleoid condensation via chloramphenicol treatment had negligible effects on HU dynamics and localizations.

We next tested the effect of nucleoid expansion on HU dynamics through rifampicin treatment (200 µg/mL, 15 min). Rifampicin is a global transcription initiation inhibitor[102] and leads to rapid degradation of unstable RNAs[103,104]. We verified that RNA production was halted (not shown) and that cells had similar cell lengths (Figure 3.1). Under expanded nucleoid conditions, we observed that

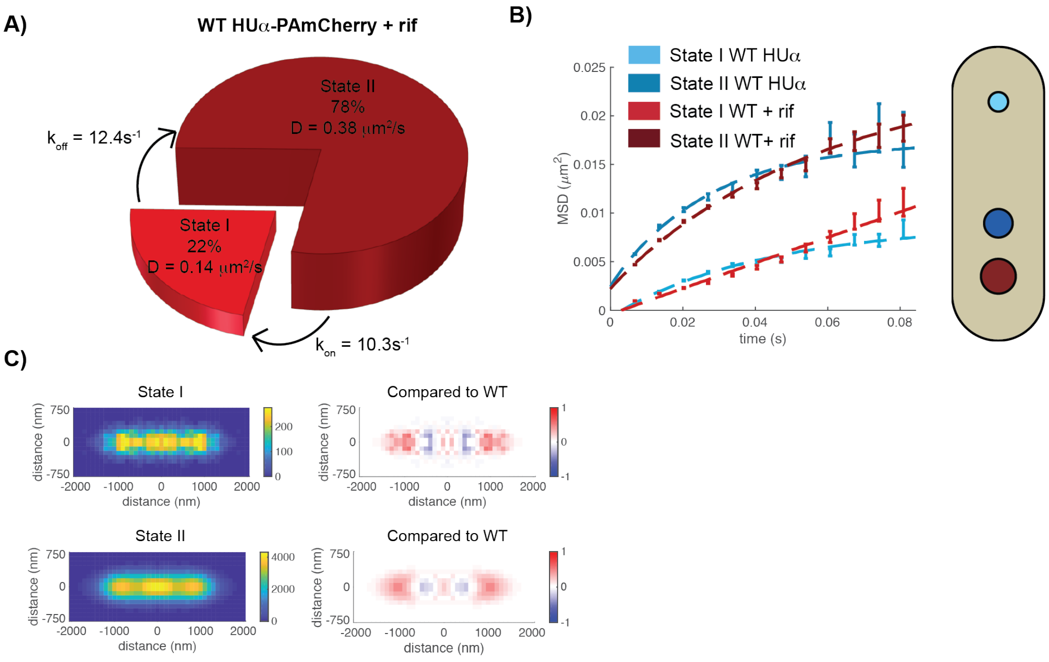


Figure 3.12: expansion of the nucleoid by rifampicin treatment shows moderate changes in HUα-PAmCherry dynamics.

(A) Two diffusive states for HUα-PAmCherry molecules under rifampicin treatment with their respective diffusion coefficients, transition rates, and population percentages identified from the HMM. (B) MSD values for HUα-PAmCherry under rifampicin treatment (light and dark red) as compared to untreated cells (light and dark blue) for State I and State II. Representative confinement sizes relative to a typical cell are demonstrated to the right. (C) 2D histograms of HUα-PAmCherry localizations under rifampicin treatment for State I and II within a normalized 4 µm x 1.5 µm cell. The bin size was 100 x 100 nm. A total of 7,925 and 129,094 localizations were used histogram construction. Comparisons to the untreated condition are shown in the right-hand column. Values represent log2-fold change in normalized density.

HUα-PAmCherry diffused nearly identically to untreated cells (n = 137,019 displacements from 87 cells), but the population percentage of bound State I molecules was reduced to 22% compared to 45% in untreated cells (Figure 3.12A). The apparent pseudo on-rate k_{on} was also reduced ~ 35% to 10.3 s^{-1}, and k_{off} increased to 12.4 s^{-1}, leading to a decreased dwell time in State I and subsequent increased dwell time in State II (Figure 3.12A). The presence of two states with similar apparent diffusion coefficients but altered kinetics suggested

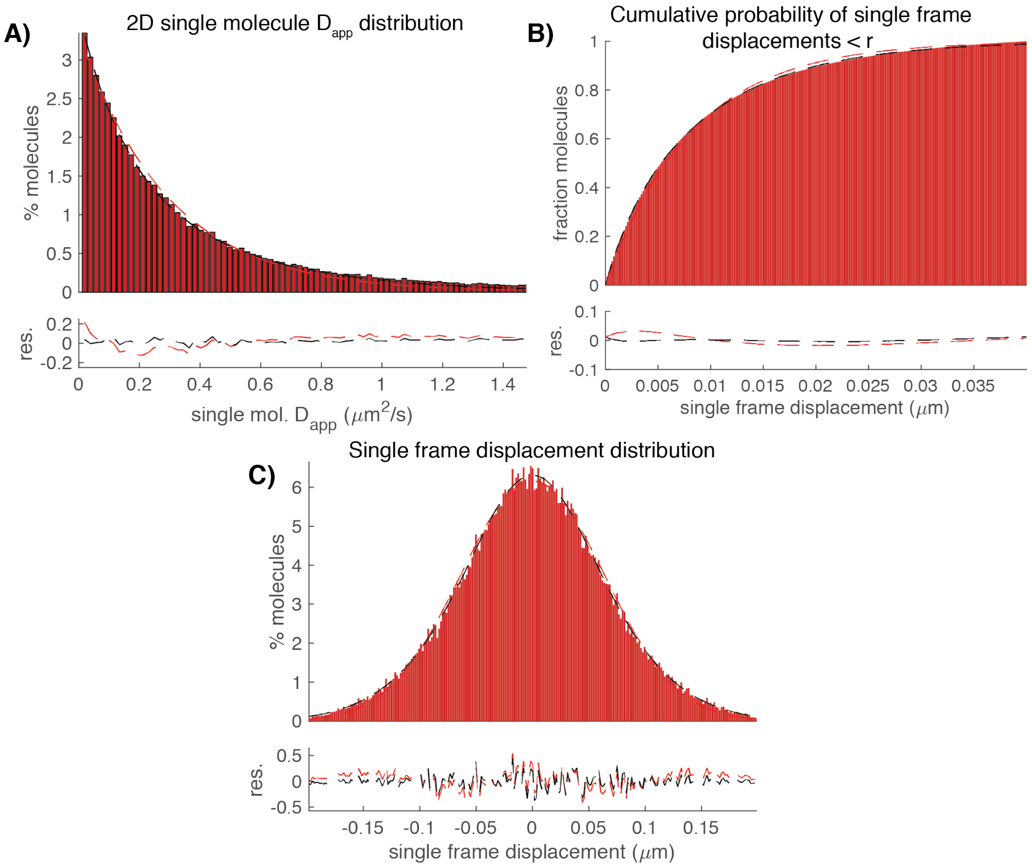


Figure 3.13: different methods to determine the number, diffusion coefficients, and occupation percentages of states for HUα-PAmCherry with rifampicin treatment.

A) For each individual displacement, the diffusion coefficient was calculated in two dimensions and binned. (B) Single frame displacements in two dimensions were calculated and the cumulative probability of displacements less than or equal to a value *r* was determined. (C) Single frame displacements were calculated along the long x-axis of the cell and binned. Red lines indicate single population fits, while black lines indicate two population fits. Residuals for each fit are plotted below each distribution. Equations used to fit each distribution are listed in Section 5. Fit parameters determined from the fits are listed in Table 3.7.

that HUα-PAmCherry molecules diffused similarly in the nucleoid, but with weaker and more transitory interactions with chromosomal DNAs compared to that in untreated cells. Consistent with this possibility, the MSD plot of HUα-PAmCherry molecules in State I followed a straight line as what would be expected in random diffusion (Figure 3.12B), indicating that the interaction of HU

Table 3.7: fit parameters from the various classical methods of determining the diffusive states represented in SPT data for HUα-PAmCherry cells treated with rifampicin.

The methodologies are 1D SFD, i.e. one dimensional single frame displacement distribution from Figure C;, the CDF of the cumulative displacement probability function from Figure B, and 2D ADD, i.e. the two-dimensional apparent diffusion coefficient distribution from Figure A. Values listed are means ± 95% confidence intervals.

HUα-PAmCherry cells treated with rifampicin				
Methodology	D_1 ($\mu m^2/s$)	p_1 (%)	D_2 ($\mu m^2/s$)	p_2 (%)
1D SFD	0.15 ± 0.01	28 ± 5	0.42 ± 0.17	72 ± 5
CDF	0.12 ± 0.02	14 ± 2	0.35 ± 0.006	86 ± 2
2D ADD	0.14 ± 0.06	22 ± 2	0.38 ± 0.07	78 ± 2

with DNA in the absence of RNAs indeed was much less constrained at a similar time scale. State II molecules remained confined, with a confinement zone moderately larger than WT State II molecules (Figure 3.12B, compare dark red with dark blue). Interestingly, the cellular distributions of HUα-PAmCherry molecules of both states displayed a three-lobed pattern with one additional lobe in the mid-cell, substantially different from that of untreated cells (Figure 3.12C). Importantly, this localization pattern also deviated significantly from the smooth, single-lobed nucleoid structure in rifampicin-treated cells[59,105,106], suggesting that in the absence of cellular RNAs, HU may interact with different chromosomal DNA sites at specific cellular localizations.

Because HU has been shown to bind many species of RNAs, including tRNA, mRNA, and specifically a non-coding RNA from the REP325 locus[85-88], we reasoned the moderately altered dynamics and localizations in rifampicin treatment could be due to the loss of these RNA binding partners, rather than a

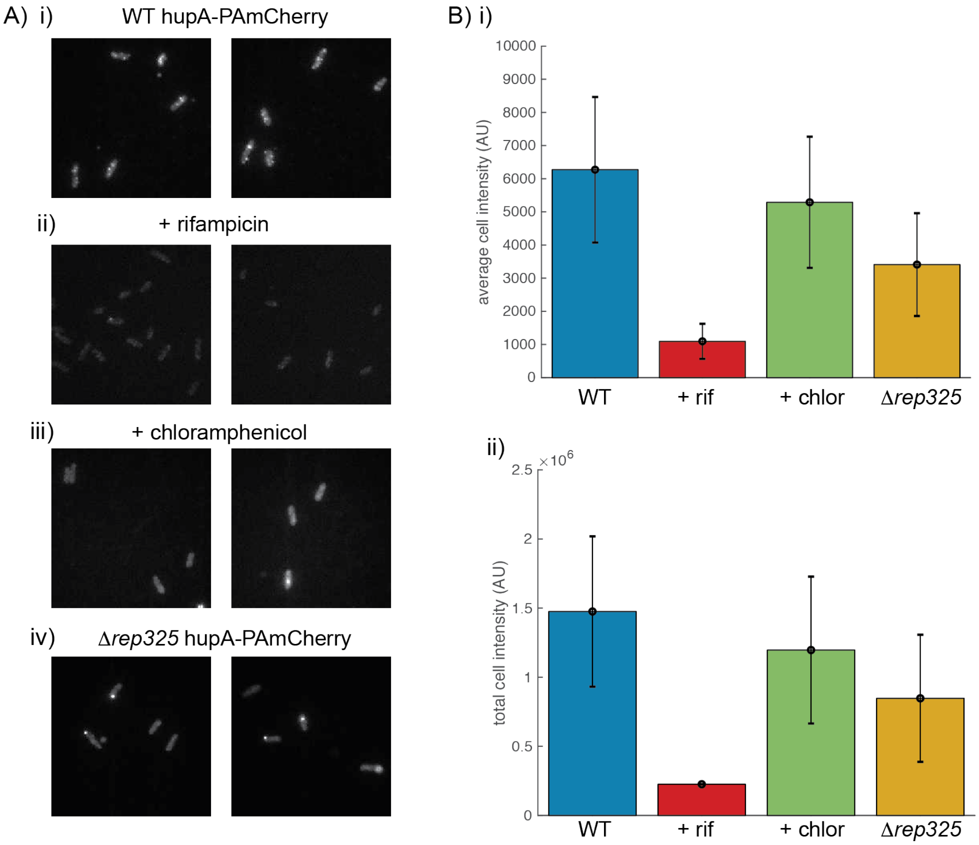


Figure 3.14: RNA FISH against the *REP325* ncRNA shows diminished levels in Δ*REP325* and rifampicin-treated cells.

(A) Example FISH images of (i) hupA-PAmCherry, (ii) hupA-PAmCherry under rifampicin treatment, (iii) hupA-PAmCherry under chloramphenicol treatment, and (iv) *ΔREP325* hupA-PAmCherry. (B) Quantification of (i) average cell intensity and (ii) total integrated cell intensity for the conditions in (A).

change in nucleoid organization. The *REP325* element produces six highly homologous small ncRNAs (naRNA1 to 6 for nucleoid-associated ncRNA), which were shown to bind to HU and condense relaxed plasmid DNA *in vitro*[87,88], implicating their role in HU-mediated chromatin organization. To determine if the altered dynamics in rifampicin treatment were indeed due to loss of these specific RNA binding partners, we constructed a strain in which the *REP325* sequence was deleted. We then integrated the *hupA-PAmcherry* fusion gene into the

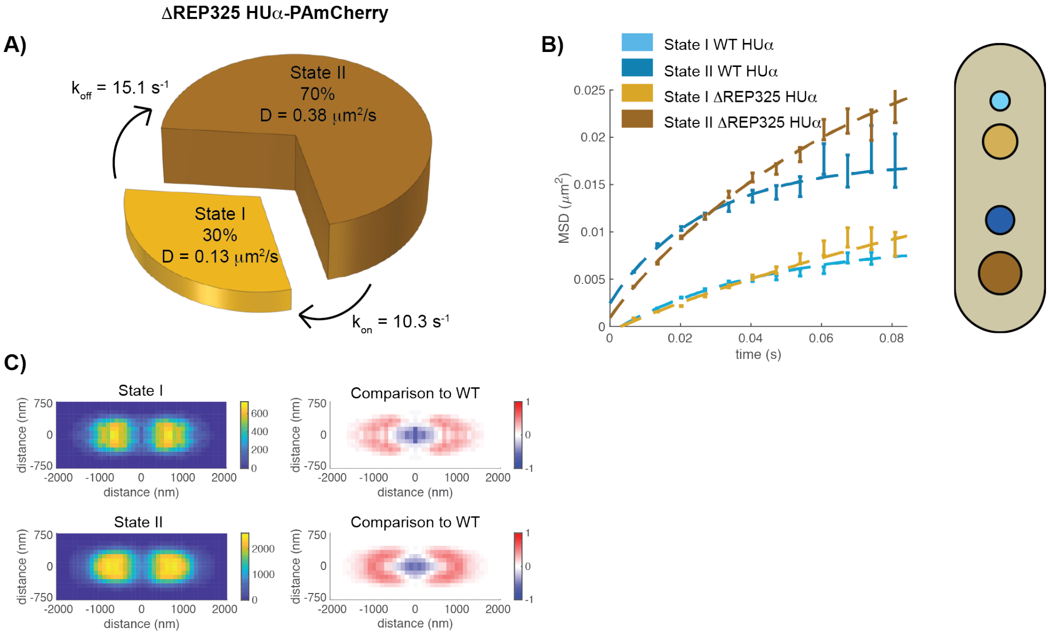


Figure 3.15: dynamics of HUα-PAmCherry in a ΔREP325 background have similar dynamics to rifampicin-treated cells.

(A) Two diffusive states of HUα-PAmCherry in the Δ*REP325* background with their respective diffusion coefficients, transition rates, and population percentages identified in the HMM. (B) MSD values of HUα-PAmCherry in the *ΔREP325* background (light and dark gold) as compared to the WT strain (light and dark blue) for State I and II. Representative confinement sizes within a normalized cell are shown to the right. (C) 2D histograms of HUα-PAmCherry localizations in the Δ*REP325* background for State I and II molecules in a normalized 4 μm x 1.5 μm cell. The bin size was 100 x 100 nm. A total of 23,003 and 91,182 for State I and State II respectively were used for the histograms. Comparison to the WT condition is shown in the right-hand column. Values represent the log2-fold change in normalized density.

Δ*REP325* strain replacing the endogenous *hupA* gene, confirmed the absence of the RNAs using FISH (Figure 3.14), and again tracked the dynamics of HUα-PAmCherry. We found that HUα-PAmCherry molecules in the Δ*REP325* the corresponding off rate increased ~ 30% to 13.8 s^{-1}. The on-rate from State II to State I similarly decreased ~15% to 12.9 s^{-1} (Figure 3.15A, Table 1). Unlike rifampicin-treated cells, State I molecules in the Δ*REP325* strain still experienced confinement, though with a larger boundary than WT (Figure 3.15B).

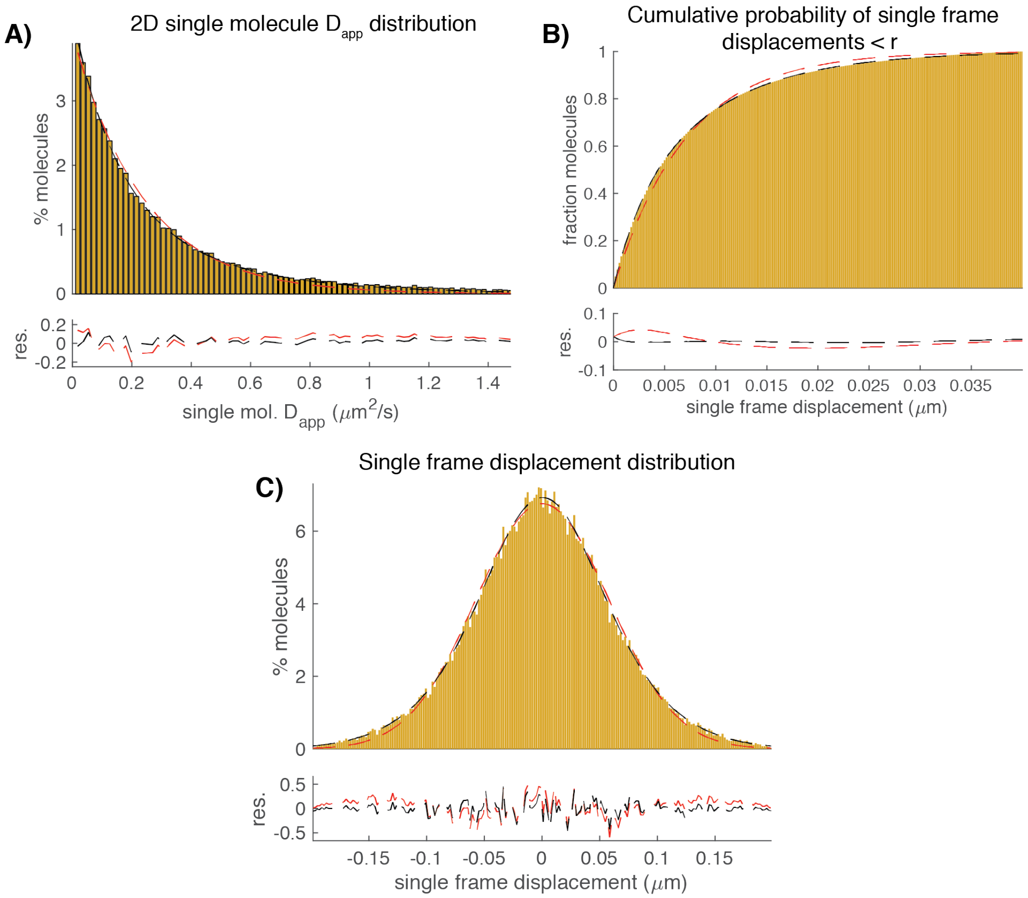


Figure 3.16: different methods to determine the number, diffusion coefficients, and occupation percentages of states for HUα-PAmCherry in a Δ*REP325*

A) For each individual displacement, the diffusion coefficient was calculated in two dimensions and binned. (B) Single frame displacements in two dimensions were calculated and the cumulative probability of displacements less than or equal to a value *r* was determined. (C) Single frame displacements were calculated along the long x-axis of the cell and binned. Red lines indicate single population fits, while black lines indicate two population fits. Residuals for each fit are plotted below each distribution. Equations used to fit each distribution are listed in Section 5. Fit parameters determined from the fits are listed in Table 3.8.

Curiously, however, cellular distributions of HUα-PAmCherry in Δ*REP325* cells exhibited a distinct, two-lobed pattern with a clear cleft at mid-cell, but seemed to avoid the midcell and prefer to localize more towards the cell pole (Figure 3.15). Interestingly, while rifampicin-treated cells demonstrated expanded

Table 3.8: fit parameters from the various classical methods of determining the diffusive states represented in SPT data for HUα-PAmCherry cells in a ΔREP325 background.

The methodologies are 1D SFD, i.e. one dimensional single frame displacement distribution from Figure C;, the CDF of the cumulative displacement probability function from Figure B, and 2D ADD, i.e. the two-dimensional apparent diffusion coefficient distribution from Figure A. Values listed are means ± 95% confidence intervals.

ΔREP325 HUα-PAmCherry cells				
Methodology	D_1 ($\mu m^2/s$)	p_1 (%)	D_2 ($\mu m^2/s$)	p_2 (%)
1D SFD	0.14 ± 0.05	38 ± 7	0.39 ± 0.12	62 ± 13
CDF	0.10 ± 0.01	27 ± 3	0.33 ± 0.01	73 ± 3
2D ADD	0.13 ± 0.05	29 ± 0.2	0.33 ± 0.07	71 ± 0

nucleoids, the nucleoids of ΔREP325 cells were similar to WT cells. However, rifampicin treatment showed degradation of the *REP325* RNAs (Figure 3.14), suggesting that loss of these specific naRNAs contributes to the change in HU dynamics, but not nucleoid expansion. This suggests the loss of naRNAs can moderately contribute to the stability of HU binding to chromosomal DNA, likely through its ability to form naRNA:HU complexes. However, we note this effect is much less dramatic than either the non-specific DNA binding mutant or the structure-specific DNA binding mutant of HU, implicating a minor role of RNA-binding in HU-mediated nucleoid organization.

F. HUαα homodimers have differential dynamics as HUαβ heterodimers

Thus far, we have discerned that the DNA-binding ability of HU likely contributes more to nucleoid organization and HU dynamics than its RNA-binding ability. In WT cells, HU is able to form both HUαα homodimers and HUαβ heterodimers, with HUββ dimers nearly undetectable[75]. In our above

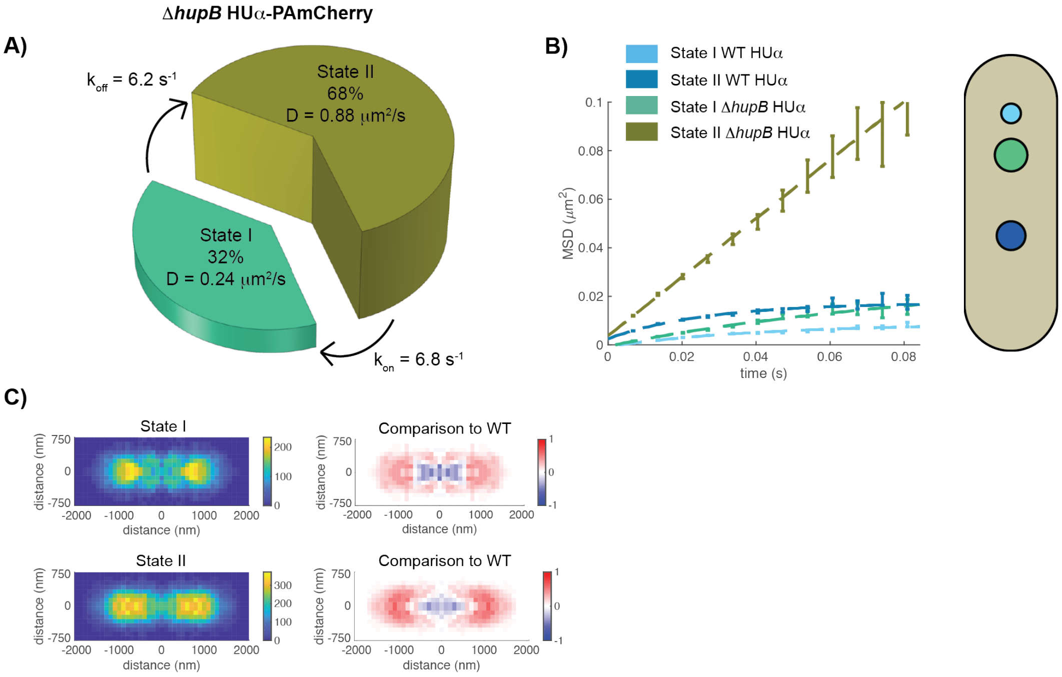


Figure 3.17: HUα-PAmCherry homodimers exhibited differential dynamics and localizations.

(A) Two diffusive states of HUα-PAmCherry in a *ΔhupB* background with their respective diffusion coefficients, transition rates, and population percentages identified from the HMM. (B) MSD values for State I and II HUα-PAmCherry in the Δ*hupB* background (mint and olive) as compared to the WT condition (light and dark blue). Representative confinement areas are shown within a normalized cell to the right. (C) 2D histograms of State I and II HUα-PAmCherry localizations in the Δ*hupB* background normalized 4 µm x 1.5 µm cell. The bin size was 100 x 100 nm. A total of 8,031 and 14,418 localizations for State I and II respectively were used for histogram construction. Comparison to the WT condition is shown in the right-hand column. Values represent the log2-fold change in normalized density.

experiments, we are unable to determine whether the observed dynamics in the DNA mutant strains are due to HUαα homodimers, HUαβ heterodimers, or a mix of the two. Because of the differential expression of the HU subunits during the growth cycle, it is implied that HUαα homodimers and HUαβ heterodimers have differential regulatory activities and/or effects on nucleoid organization. To better understand the differential roles of HU dimers, we performed SMT on HUα-PAmCherry molecules in a *ΔhupB* background, ensuring that we were only

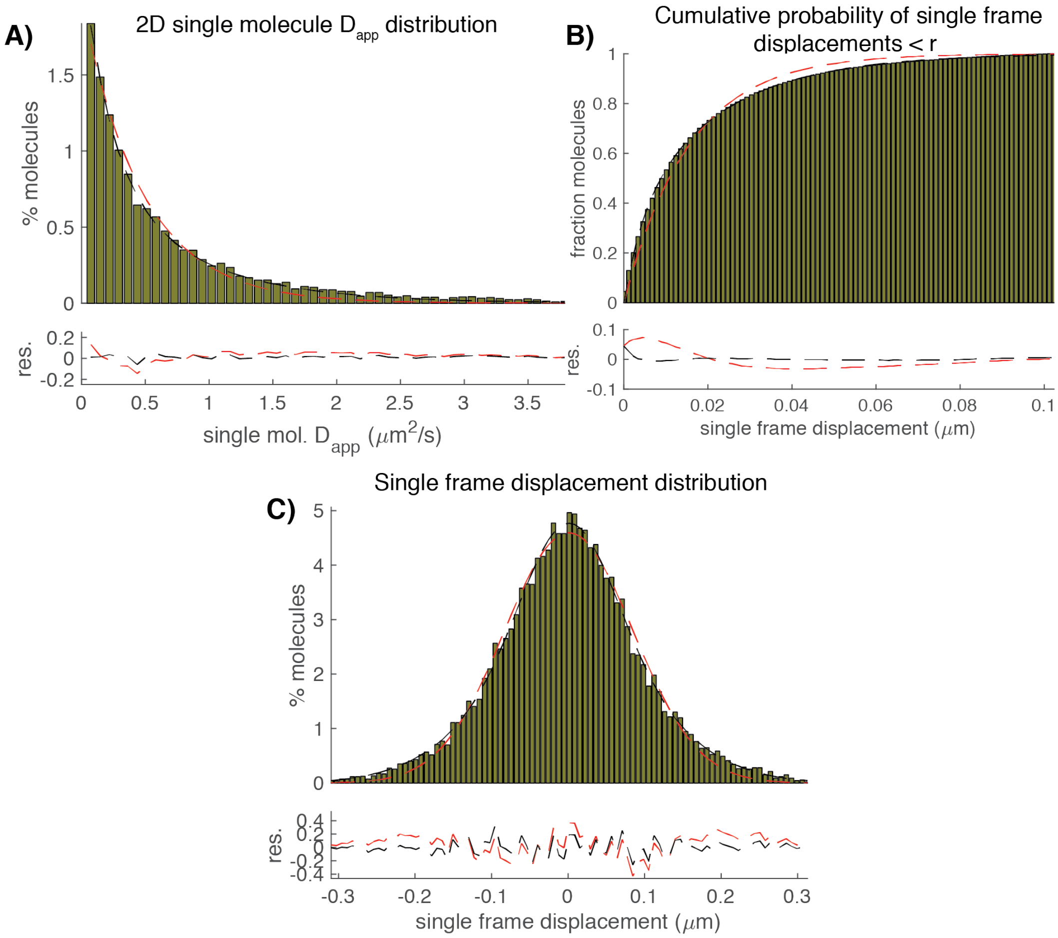


Figure 3.18: classical methods to determine the number, diffusion coefficients, and occupation percentages of states for Δ*hupB* HUα-PAmCherry

A) For each individual displacement, the diffusion coefficient was calculated in two dimensions and binned. (B) Single frame displacements in two dimensions were calculated and the cumulative probability of displacements less than or equal to a value *r* was determined. (C) Single frame displacements were calculated along the long x-axis of the cell and binned. Red lines indicate single population fits, while black lines indicate two population fits. Residuals for each fit are plotted below each distribution. Equations used to fit each distribution are listed in Section V. Fit parameters determined from the fits are listed in Table 3.9.

visualizing HUαα homodimers. Perhaps unsurprisingly, we found the dynamics of *ΔhupB* HUα-PAmCherry molecules to differ from the WT condition. Again, there were two diffusive states present in the data, with diffusion coefficients faster than the WT condition: State I with D = 0.23 $\mu m^2/s$ representing 38% of molecules and State II with D = 0.88 $\mu m^2/s$ representing 62% of molecules (n =

Table 3.9: fit parameters from the various classical methods of determining the diffusive states represented in SPT data for HUα-PAmCherry cells in a Δ*hupB* background.

The methodologies are 1D SFD, i.e. one dimensional single frame displacement distribution from Figure C;, the CDF of the cumulative displacement probability function from Figure B, and 2D ADD, i.e. the two-dimensional apparent diffusion coefficient distribution from Figure A. Values listed are means ± 95% confidence intervals.

ΔhupB HUα-PAmCherry cells				
Methodology	D_1 ($\mu m^2/s$)	p_1 (%)	D_2 ($\mu m^2/s$)	p_2 (%)
1D SFD	0.26 ± 0.13	38 ± 6	0.92 ± 0.37	63 ± 9
CDF	0.17 ± 0.04	32 ± 5	0.81 ± 0.05	68 ± 5
2D ADD	0.23 ± 0.09	32 ± 1	0.77 ± 0.17	68 ± 1

22,449 displacements from 48 cells) (Figure 3.17A). Like the WT condition, HUαα homodimers exhibited transitory dynamics, with slightly longer dwell times of 168 ms and 155 ms for States I and II respectively. Concordant with their increased diffusive behavior, the confinement of State I increased by > two-fold to ~ 450nm, while State II diffused with normal Brownian motion (Figure 3.17B). The localizations of HU were significantly depleted in the mid-cell with localizations peaking at the polar edge of the nucleoid, opposite of WT cells (Figure 3.17C). Intriguingly, the dynamics and localizations of HUαα homodimers were extraordinarily similar to HUα(P63A)-PAmCherry in the presence of HUβ. This suggests that HUα(P63A) may not bind HUβ well, that HUαβ needs both subunit prolines to function, and/or that HUββ homodimers preferentially form when HUα is inactivated in this manner. Regardless, these results suggest that HUαα homodimers indeed have differential dynamics to HUαβ heterodimers within the cell.

G. HUαα homodimers play a role in nonspecific chromatin organization but not site-specific activity

HUα(triKA)-PAmCherry showed a dramatic effect in terms of dynamics of HU, although the cells had normal nucleoids and growth rates, presumably due to compensation by the HUβ subunit. To investigate the effect of the triple lysine mutation on HUαα homodimers alone, we imaged HUα(triKA)-PAmCherry in the *ΔhupB* background strain. Without HUβ present, HUα(triKA)-PAmCherry homodimers showed overall increased rates of diffusion with two diffusive states of 0.4 ± 0.01 $\mu m^2/s$ and 2.42 ± 0.03 $\mu m^2/s$ with occupations of 42 ± 1% and 58 ± 1% respectively (μ ± std, n = 58,867 displacements from 68 cells) (Figure 3.19A). Concomitantly, both states had greatly increased confinement areas, with State I confinement at 630 ± 1 nm in diameter, and State II having a confinement size of 1340 ± 1 nm in diameter (μ ± 95% confidence interval, Figure 3.19B).

Additionally, localizations of HUα(triKA)-PAmCherry homodimers were more expanded along the x-axis compared to HUα-PAmCherry homodimers (Figure 3.19C). Interestingly, compared to HUα(triKA)-PAmCherry, Δ*hupB* HUα(triKA)-PAmCherry homodimers displayed a more narrow distribution along the x- and y-axes (Figure 3.19C), implying that without HUβ, HUα(triKA)-PAmCherry regains partial function, at least in terms of proper localization to the nucleoid. However, we find that Δ*hupB* HUα(triKA)-PAmCherry cells were

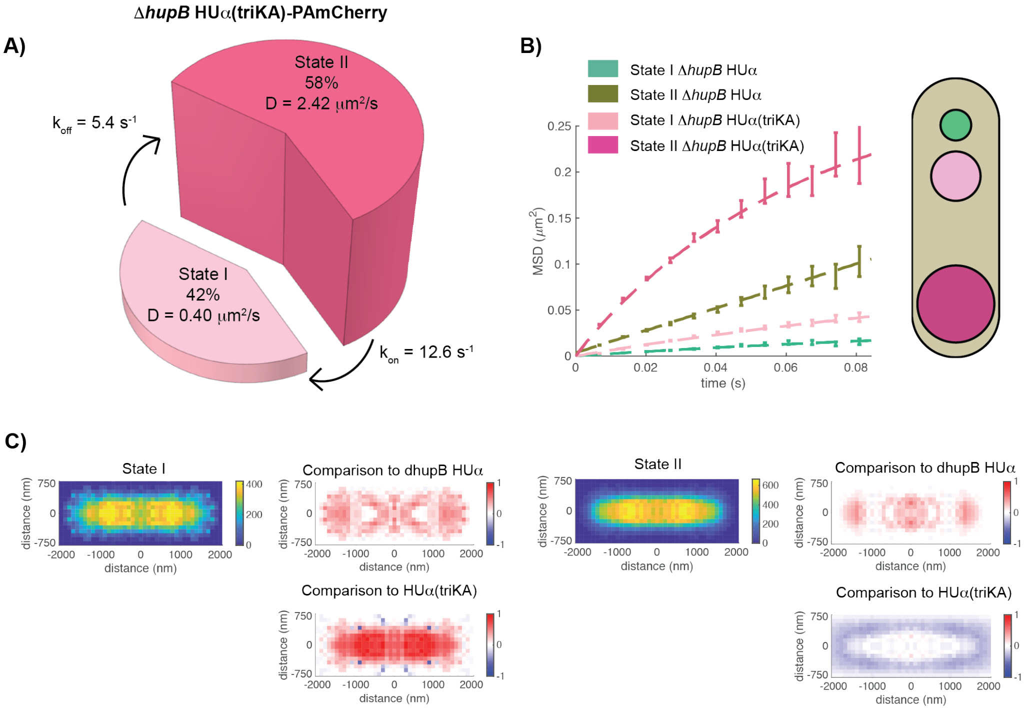


Figure 3.19: non-specific DNA binding mutations in HUα-PAmCherry homodimers exhibit altered dynamics but regain partial nucleoid localization.

(A) Two diffusive states for HUα(triKA)-PAmCherry in a Δ*hupB* background with their respective diffusion coefficients, transition rates, and population percentages as identified by the HMM. (B) MSD values for HUα(triKA)-PAmCherry in a Δ*hupB* background (light and dark pink) as compared to the WT strain (light and dark blue) for State I and II molecules. (C) 2D histograms of State I and II HUα(triKA)-PAmCherry molecules in the Δ*hupB* background within a normalized 4 µm x 1.5 µm cell. The bin size was 100 x 100 nm. A total of 24,695 and 34,172 localizations were used for histogram reconstructions. Comparison to the WT HUα-PAmCherry in the ΔhupB strain and to the HUα(triKA)-PAmCherry strain with HUβ present is provided in the right-hand column. Values represent the log2-fold change in normalized density compared to the two conditions.

significantly longer than WT cells (Figure 3.7) which may overestimate the extent of distribution changes. The deviations in cell length suggest this strain may have replication and/or cell division defects, hallmarks of *hupA hupB* deletion[107] (Figure 3.23), implying this mutant is as detrimental as total loss of the HU protein complex. Interestingly, the the *ΔhupB hupA(triKA)-PAmCherry* strain displayed more elongated cells than the Δ*hupA* Δ*hupB* strain (Figure 3.7 and Figure 3.23) and grew significantly more slowly

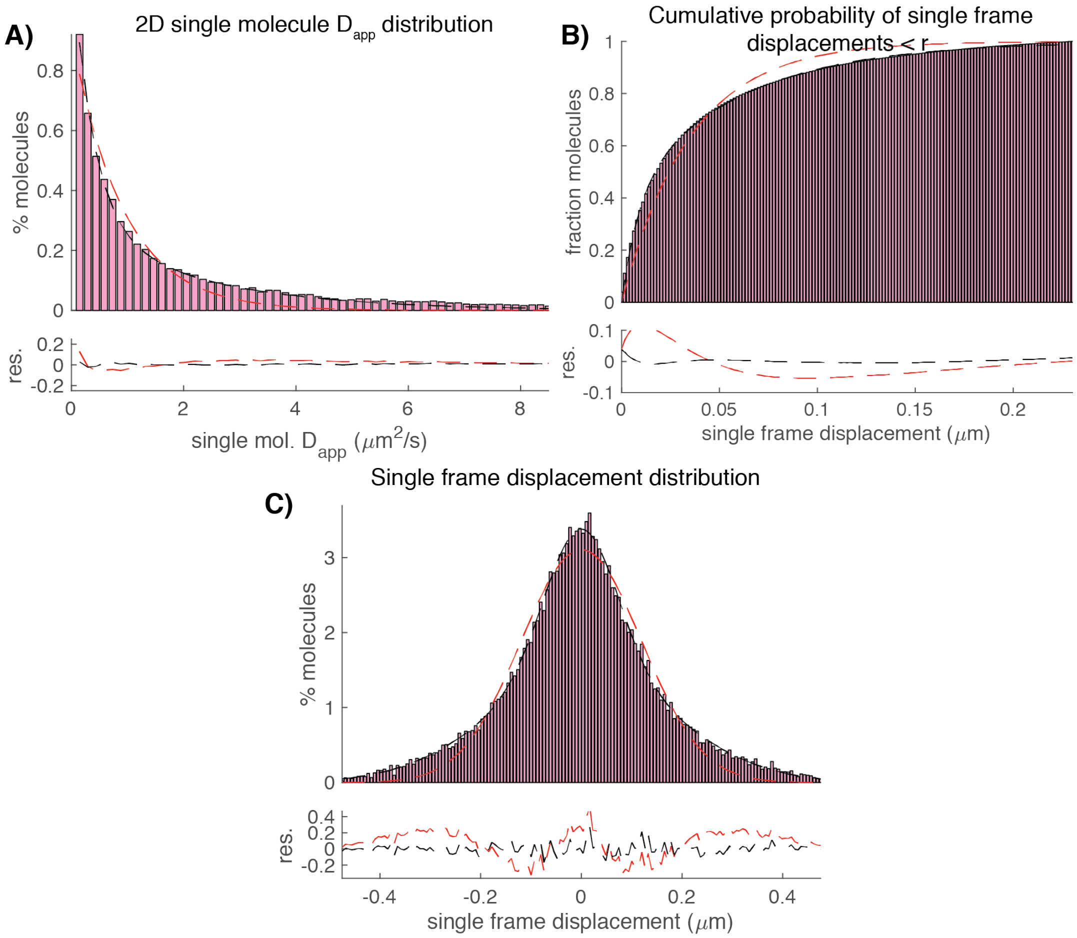


Figure 3.20: classical methods for determining the number, diffusion coefficients, and occupation percentages of Δ*hupB* HUα(triKA)-PAmCherry.

A) For each individual displacement, the diffusion coefficient was calculated in two dimensions and binned. (B) Single frame displacements in two dimensions were calculated and the cumulative probability of displacements less than or equal to a value *r* was determined. (C) Single frame displacements were calculated along the long x-axis of the cell and binned. Red lines indicate single population fits, while black lines indicate two population fits. Residuals for each fit are plotted below each distribution. Equations used to fit each distribution are listed in Section 5. Fit parameters determined from the fits are listed in Table 3.10.

than the Δ*hupA* Δ*hupB* strain (Figure 3.1B), signifying this mutant may be more harmful than full HU subunit deletion. This worsening of cell fitness is curious given the state-switching dynamics and localizations appear to be more WT-like than HUα(triKA)-PAmCherry with WT HUβ. Previously, we found that HUα(P63A)-PAmCherry molecules had increased diffusion coefficients and

Table 3.10: fit parameters from the various classical methods of determining the diffusive states represented in SPT data for HUα(triKA)-PAmCherry cells in a *ΔhupB* background.

The methodologies are 1D SFD, i.e. one dimensional single frame displacement distribution from Figure C;, the CDF of the cumulative displacement probability function from Figure B, and 2D ADD, i.e. the two-dimensional apparent diffusion coefficient distribution from Figure A. Values listed are means ± 95% confidence intervals.

ΔhupB HUα(triKA)-PAmCherry cells				
Methodology	D_1 ($\mu m^2/s$)	p_1 (%)	D_2 ($\mu m^2/s$)	p_2 (%)
1D SFD	0.33 ± 0.03	40 ± 3	2.2 ± 0.08	59 ± 3
CDF	0.38 ± 0.07	35 ± 8	2.5 ± 0.04	65 ± 16
2D ADD	0.38 ± 0.8	35 ± 0.01	2.23 ± 0.4	65 ± 0.1

confinement sizes, exhibiting diffusive behavior surprisingly similar to HUαα homodimers, suggesting that HUβ is either rendered non-functional or HUαα dimers preferentially form under this mutation. If accurate, HUα(P63A)-PAmCherry homodimers should display similar diffusive behavior, dynamics, and localizations to HUα(P63A)-PAmCherry in the presence of HUβ as well as the WT HUαα homodimers. To investigate the validity of this hypothesis, we inserted the *hupA(P63A)-PAmCherry* fragment into the Δ*hupB* strain. Without HUβ, we observed two diffusive states of HUα(P63A)-PAmCherry: State I with diffusion coefficient of 0.29 ± 0.01 $\mu m^2/s$, and State II

with diffusion coefficient of 1.02 ± 0.01 $\mu m^2/s$, with occupations of 30 ± 1% and 70 ± 1% respectively (μ ± s.e.m., n = 132,631 displacements from 67 cells) (Figure 3.21A).

The kinetics of state switching remained highly dynamic, with dwell times

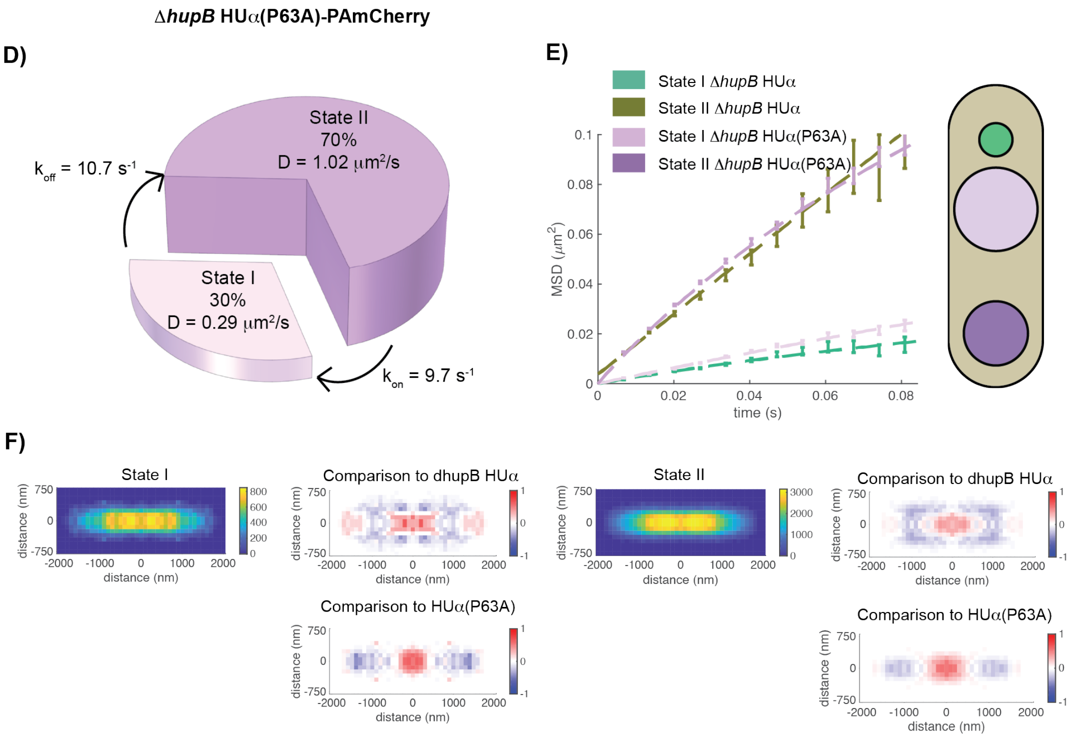


Figure 3.21: Structure-specific binding mutation in HUα-PAmCherry homodimers has minor effect on dynamics and localization compared to WT HUα-PAmCherry homodimers.

(A) Two diffusive states for HUα(P63A)-PAmCherry in a Δ*hupB* background with their respective diffusion coefficients, transition rates, and population percentages as identified by the HMM. (B) MSD values for the HUα(P63A)-PAmCherry molecules in a Δ*hupB* background (light and dark purple) as compared to the WT condition (light and dark blue) for State I and II. Representative confinement areas are shown in a normalized cell to the right. (C) 2D histograms of State I and II HUα(P63A)-PAmCherry localizations in Δ*hupB* background strain within a normalized 4 x 1.5 µm cell. The bin size was 100 x 100 nm. A total of 27,287 and 105,324 localizations were used for histogram reconstructions. Comparison to the WT ΔhupB HUα-PAmCherry and HUα(P63A)-PAmCherry strains is located in the right-hand column. Values represent the log2-fold change in normalized density compared to the two conditions.

of 100 ± 8 ms and 110 ± 6 ms for States I and II respectively, in the realm of dwell times for both the P63A molecules with HUβ present and the WT HUαα homodimers. Despite similar dynamics to the *ΔhupB* HUα-PAmCherry strain, the confinement values for States I and II in the Δ*hupB* HUα(P63A) strain were rather large, 849 ± 4 nm and 1010 ± 1 nm respectively (µ ± 95% CI) (Figure 3.21B).

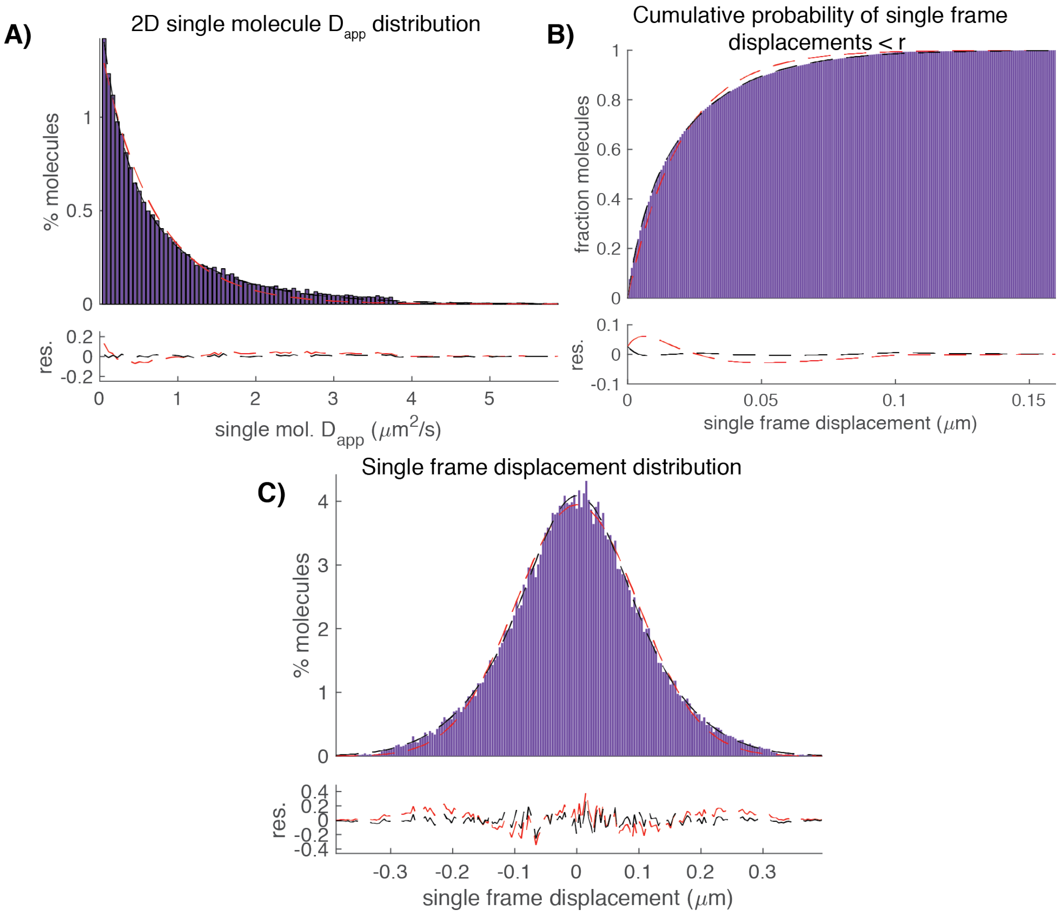


Figure 3.22: classical methods to determine diffusion coefficients and state percentages for HUα(P63A)-PAmCherry homodimers.

A) For each individual displacement, the diffusion coefficient was calculated in two dimensions and binned. (B) Single frame displacements in two dimensions were calculated and the cumulative probability of displacements less than or equal to a value *r* was determined. (C) Single frame displacements were calculated along the long x-axis of the cell and binned. Red lines indicate single population fits, while black lines indicate two population fits. Residuals for each fit are plotted below each distribution. Equations used to fit each distribution are listed in Section 5. Fit parameters determined from the fits are listed in Table 3.11.

Furthermore, Δ*hupB* HUα(P63A)-PAmCherry molecules localized more towards the center of the cell, similarly to the WT condition (Figure 3.21). While relatively unchanged in their behavior to ΔhupB HUα-PAmCherry and HUα(P63A)-PAmCherry in the presence of HUβ*,* their phenotype was more severe, demonstrating significantly elongated cells (Figure 3.7 and Figure 3.23) and

Table 3.11: fit parameters from the various classical methods of determining the diffusive states represented in SPT data for HUα(P63A)-PAmCherry cells in a *ΔhupB* background.

The methodologies are 1D SFD, i.e. one dimensional single frame displacement distribution from Figure C;, the CDF of the cumulative displacement probability function from Figure B, and 2D ADD, i.e. the two-dimensional apparent diffusion coefficient distribution from Figure A. Values listed are means ± 95% confidence intervals.

ΔhupB HUα-PAmCherry cells				
Methodology	D1 (µm2/s)	p1 (%)	D2 (µm2/s)	p2 (%)
1D SFD	0.2 ± 0.02	25 ± 2	0.94 ± 0.002	75 ± 2
CDF	0.3 ± 0.1	29 ± 2	1.1 ± 0.2	71 ± 5
2D ADD	0.29 ± 0.06	25 ± 0.01	1.02 ± 0.07	75 ± 0.01

doubling times (Table 3.2). We note, however, that unlike the Δ*hupB* HUα(triKA)-PAmCherry strain, the Δ*hupB* HUα(P63A)-PAmCherry cells have normal physiology at 37°C (Figure 3.23). Because the P63A mutation abolishes structure-specific binding at regulatory sequences, this may indirectly imply the role of HU in regulating cold-shock genes. Curiously, while the diffusive and dynamic behavior of HUα(P63A)-PAmCherry homodimers was indeed similar to WT HUα-PAmCherry homodimers and HUα(P63A)-PAmCherry in the presence of HUβ, the localizations more resembled WT cells. This observation notwithstanding, this data ultimately suggests the site-specific activity of HU may require HUαβ heterodimers, or at least HUβ present within the cell.

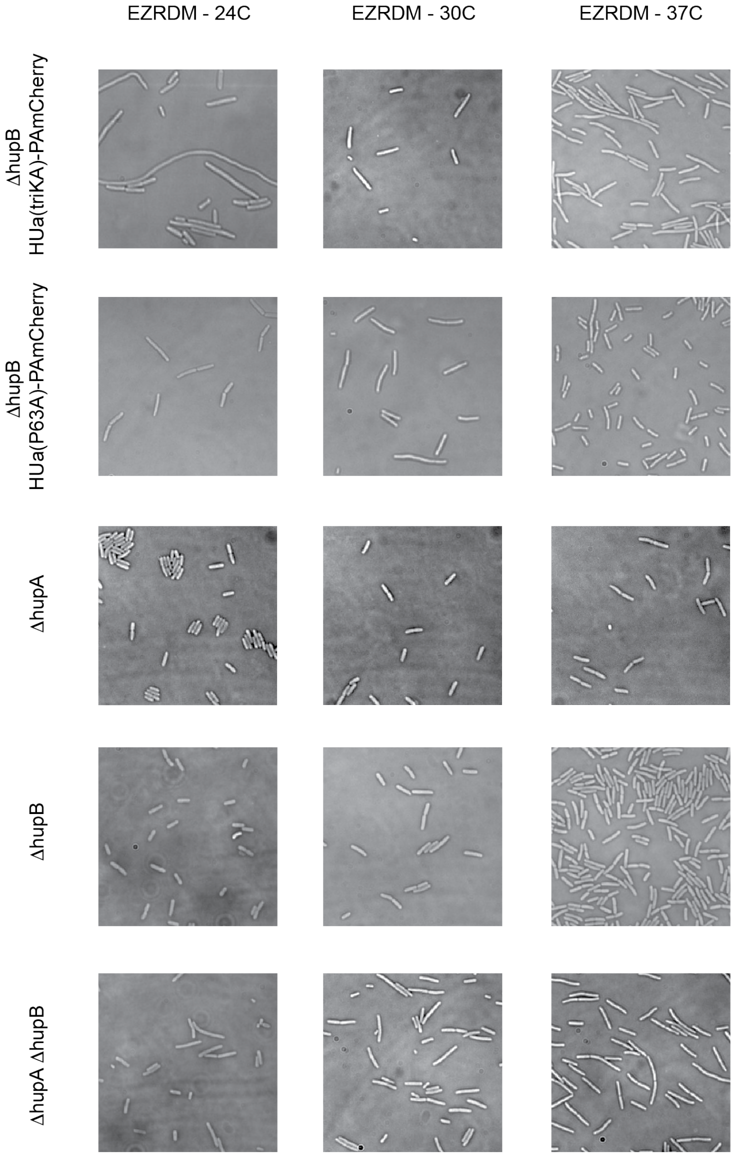


Figure 3.23: brightfield images of HU subunit deletions compared to DNA-binding mutants.

Comparison of cell phenotypes in *ΔhupB hupA(triKA-PAmCherry*, *ΔhupB hupA(P63A)-PAmCherry*, *ΔhupA*, *ΔhupB*, and *ΔhupA ΔhupB* strains at three temperatures, 24°C, 30°C, and 37°C.

III. Discussion

Using superresolution single molecule tracking of HUα-PAmCherry molecules in live cells, we demonstrate that HUα-PAmCherry has two diffusive states with diffusion coefficients of ~0.1 and ~0.4 $\mu m^2/s$ for State I and State II respectively, corresponding to DNA-bound and nucleoid-diffusing HUα-PAmCherry states, respectively. We found that HU was highly dynamic and that it switches rapidly between the two states. Assuming an average concentration of nonspecific chromosomal DNA binding sites at ~ 4 mM under our growth condition and based on the two transition rates k_{on} and k_{off}, we estimated that the nonspecific binding affinity of HUα-PAmCherry to chromosomal DNA was at ~ 5 mM, and the association rate constant was at ~1 x 10^3 $M^{-1}s^{-1}$, both rates two to three orders of magnitude lower than those specific DNA or RNA binding proteins [108,109]. We found that HU localized strongly to the nucleoid region of the cell, with high density near the mid-cell edge of the two nucleoids normally present in fast-growing cells (Figure 3.2D). Interestingly, a recent study found that in stationary phase cells, a gradient of negative supercoiling exists from the origin to the terminus, and that this increase in negative supercoiling requires HUβ[92]. In our growth conditions, it is plausible to have a substantial fraction of HUβ present in the cell, which may explain the more mid-cell localized distribution of HUα-PAmCherry. The transient dynamics of HU were surprising as many models of nucleoid organization suggest stably bound HU dimers and/or multimers on the DNA, where the coordinated action of multiple HU dimers can stabilize their interactions with DNA and create a network of HU dimer-dimers and DNA interactions[89,90]. If these models are relevant *in vivo*, this could suggest that instead of stable structures, these HU multimers are constantly exchanging with the cytoplasmic pool. However, given that our PALM images of HUα-PAmCherry demonstrated that HUα-PAmCherry was homogenously distributed across the nucleoid, we hypothesize this model is likely not a major contributing factor. We propose the mechanism of HU-mediated chromosomal organization is likely due to the cumulative forces of the thousands of transient HU-DNA interactions. This is reminiscent of the weak forces of crowding in protein association, where individual contributions are weak, but the cumulative effect strongly promoters protein association[30]. This is analogous to the coordinated mass action of ants to act on objects several times larger than the size of an individual ant. However, more biophysical characterization of this proposed mechanism is needed.

Given the small confinement of HU at ~ 230 and 300 nm for the stably DNA-bound State I and loosely DNA-associated State II respectively, HU may unbind and rebind at or near its previously bound site. Such dynamic turnover may facilitate binding of transcription factors and other DNA-binding proteins, and/or allow for rapid reorganization of the local chromatin structure, an ability of HU demonstrated by its gain-of-function mutant[84]. Additionally, a heterologous protein in humans, HMGB-1, has been demonstrated through FRAP experiments to be highly mobile, suggesting equally transient interactions on the chromatin as HU[110]. As the role of HMBG-1 is likewise to modulate local chromatin organization[83,111,112], we theorize these two proteins may mediate chromatin organization through similar mechanisms.

Unlike most other bacteria, certain enterobacteria including *E. coli* have two highly homologous subunits of HU (HUα and HUβ) instead of a single subunit. Although these two subunits have extremely high affinity for HUαβ heterodimer formation, their genes are located > 400 kbps apart on the chromosome. During lag phase, only HUα subunits are expressed, and HUβ is expressed in increasing quantities until a peak in stationary phase, where nearly all HU exists as HUαβ heterodimers[75]. Because of this differential composition during cell growth, it has long been proposed that the different dimers- HUαα and HUαβ in particular- may play differential roles in chromosome organization[90]. In our imaging condition for WT HUα-PAmCherry, we could not determine if the observed molecules were HUαα dimers, HUαβ dimers, or a mixture of the two. When we imaged *ΔhupB* HUα-PAmCherry molecules, thus ensuring we were exclusively imaging HUαα dimers, we found they exhibited dynamics significantly different than WT HUα-PAmCherry. The two diffusive states were retained, but with increased diffusion coefficients and larger confinement areas, suggesting an altered affinity for DNA compared to the WT HUα-PAmCherry strain. State-switching kinetics between the two diffusive states remained highly dynamic, indicating that both types of HU dimers have greatly transitory, non-specific interactions across the nucleoid. The apparent lower affinity of HUαα dimers for DNA is consistent with prior observations of the nucleoid during different growth phases. In lag phase, where HUαα is dominantly expressed, the chromosome is much more decondensed compared to the more condensed structures found in log-phase and stationary phase growth[90]. These observations suggest that these relatively small differences in affinity of the HU dimers can cause global changes in the organization state of the nucleoid. Interestingly, the localization of HUαα dimers differed substantially from the WT HUα-PAmCherry strain, still preferring nucleoid localization but with higher density near the polar edge of the nucleoid and lower density in mid-cell (Figure 3.17C). This is consistent with our previous hypothesis that in the WT HUα-PAmCherry strain, where HUβ likely constitutes a significant fraction of the overall pool of HU subunits, HUβ association directs HUα-PAmCherry to the terminus region[92].

While its importance is clear, the mechanism by which HU-mediated nucleoid organization is achieved has remained elusive. To investigate, we mutated residues in the HUα subunit that affect its two DNA-binding modes: (i) non-specific binding across the nucleoid, and (ii) structure-specific binding with respect to regulation of transcription. Mutating three key lysine residues on the surface of HUα important for non-specific binding to DNA dramatically altered the dynamics of HUα-PAmCherry, essentially abolishing the DNA-bound State I, and significantly increased the diffusion coefficient of State II to that of a more freely diffusing cytoplasmic protein. The high positive surface charge of HU dimers ostensibly contributes to its strong nucleoid localization; hence, mutation of these key residues likely reduces the dipole-dipole interactions that keep it anchored in the nucleoid. Consistent with this, HUα(triKA)-PAmCherry molecules localized significantly more broadly across the cell compared to WT cells (Figure 3.5C). Charge alone does not necessarily dictate protein localization, however; when the surface charge of GFP was changed to be highly positive, the modified molecules actually avoided the nucleoid region[113]. Nevertheless, the crystal structures in coordination with the phenotypic effects we observed highlight the critical nature of a positively-charged surface to non-specific DNA binding, and consequently the importance of HU non-specific binding to DNA for chromosome organization. Strikingly, the triple lysine mutant HUα in the presence of HUα displayed distinct dynamics than the triple lysine HUαα homodimer. HUα(triKA)-PAmCherry with HUβ present localized quite broadly across the cell, while the HUα(triKA) homodimer localized more similarly to WT HUαα dimers (Figure 3.19C), indicating that HUα(triKA) homodimers have more functionality in terms of proper spatial localization than HUα(triKA) with HUβ. This insinuates that HUα(triK)-PAmCherry either cannot form heterodimers and HUββ is responsible for the relatively normal nucleoid volume, or that when HUα(triKA)-PAmCherry is bound to HUβ, its binding affinity to DNA is worse than the mutant homodimer. Because the dynamics of the triple lysine mutant homodimer differ dramatically from that of HUα(triKA)-PAmCherry in the presence of HUβ, we hypothesize that HUα(triKA)-PAmCherry is sequestered by HUβ due to the high affinity between these two subunits, and the triple lysine mutation in HUα renders the heterodimer non-functional. When HUβ is not present, the triple lysine mutant homodimer can form unhindered and regains partial function to bind non-specifically to DNA. Both strains exhibited longer cell lengths that did not recover at higher temperatures (Figure 3.7 and Figure 3.23), suggesting this mutation has a dominant negative effect. Together, this data strongly supports the role of non-specific binding for chromosome organization. Additionally, the differences in dynamics between the triple lysine homodimer and the mixed homodimer and heterodimer strain suggest the HUαα homodimer is better able to bind non-specific DNA over the HUαβ heterodimer.

In addition to binding non-specifically throughout the nucleoid, HU has many other cellular roles inside the cell, including the regulation of gene expression by stabilizing repressor loops[97,81,98]. A virtually universally conserved proline residue in the beta sheet arm of each HU subunit is able to intercalate into the DNA duplex and stabilize looped DNA[99]. When we mutated this conserved proline residue, we again saw HUα(P63A)-PAmCherry to have differential dynamics to the WT strain, though less dramatic than with the triple lysine mutation. HUα(P63A)-PAmCherry displayed overall increased diffusive behavior with larger confinement areas, surprisingly similar in behavior to the HUαα homodimers alone as well as the HUα(P63A) homodimers alone. The nearly identical dynamics of HUα(P63A)-PAmCherry with the HUαα homodimers alone suggested that HUα was not as important to structure-specific binding as HUβ. Corroborating this supposition, HUα(P63A)-PAmCherry with HUβ had normal cell lengths and only a minor increase in cell doubling time, while

HUα(P63A)-PAmCherry in the Δ*hupB* background had extremely elongated cell lengths and significantly longer doubling times, suggesting that HUβ was sufficient to rescue the P63A mutation. Additionally, HUαα homodimers alone also demonstrated a minor increase in cell doubling time, but their cell lengths were actually shorter than WT, suggesting that while HUαα homodimers are able to perform structure-specific binding, they may not be as efficient as HUαβ heterodimers, leading to some minor cell phenotypes. Altogether, this supports assumptions made in the literature that HUαα and HUαβ have differential functions within the cell. We noted that HUα(P63A)-PAmCherry homodimers exhibited more "WT"-like localizations (i.e. more density near the mid-cell) over the WT HUαα homodimers (i.e. more density near the polar edges of the nucleoid). Consistent with the prior observation, we hypothesize this change in localization is due to the absence of HUβ. When HUβ is present, HUα(P63A)-PAmCherry is brought to the *ter* region through the HUβ subunit[92]. Without HUβ, HUα seems to preferentially localize more toward the *ori* region of the chromosome.

Because HU has a large role in chromosome organization that in turn affects gene regulation, it is possible that the observed effects on HU dynamics were due to altered nucleoid organization, and not necessarily by the mutations themselves. HUα-PAmCherry in the presence of the translation inhibitor chloramphenicol, which dramatically condenses the nucleoid[59,106], showed almost negligible changes compared to untreated HUα-PAmCherry molecules, indicating a null effect of nucleoid organization on HU dynamics. However, HUα-PAmCherry treated with the transcription inhibitor rifampicin demonstrated minor changes in dynamics, with lower occupation of the DNA-bound state and a corresponding change in state-switching dynamics, although the molecules diffused with identical diffusion coefficients as untreated cells. Because HU is known to bind to several species of RNAs, particularly the *REP325* naRNAs, we postulated the change in dynamics was caused by the loss of these RNA binding partners. Indeed, imaging HUα-PAmCherry in a Δ*REP325* background produced results strikingly similar to rifampicin-treated cells. Intriguingly, the nucleoid morphology of rifampicin-treated cells and the Δ*REP325* strain differed, and consequently HUα-PAmCherry exhibited differential localization distributions. This corroborated our supposition that nucleoid organization has minimal effect on HU dynamics. Given the vast changes in nucleoid structure capable depending on various growth and stress conditions, this result should not be surprising. Puzzling, although HU binds RNA with nanomolar affinity, deletion of RNAs- in bulk or the specific naRNAs- had a much less dramatic effect on HU dynamics than the DNA-binding mutations. This would suggest that the main mechanism of HU-mediated chromosome organization is arbitrated through its ability to bind DNA non-specifically. Yet, HU has been demonstrated directly *in vitro* to deposit naRNAs onto cruciform structures present in the cell. As cruciform DNA structures are a ubiquitous component of the genomes of all life[112], this likely indicates an undiscovered role of naRNAs and/or HU.

IV. Materials and Methods

A. Cell growth

BW25993::hupA-PAmCherry cells and ACL035 were inoculated from single colonies from freshly-streaked LB plates into EZ Rich Defined Media (EZRDM, Teknova) using 0.4% glucose as the carbon source with chloramphenicol antibiotic (Sigma-Aldrich C0378) added to 150 µg/mL, and carbinicillen antibiotic (Sigma-Aldrich C3416) added to 60 µg/mL respectively. Cells were grown overnight at room temperature (RT), shaking at 240 rpm. The following morning, saturated cultures were re-inoculated into fresh EZRDM and grow for several hours at RT until mid-log phase growth (OD600 ~0.4). For rifampicin drug treatment in the *hupA-PAmCherry* strain, rifampicin antibiotic (R3501) was added to a final concentration of 200 µg/mL for 15 minutes before cells were harvested for imaging (see Materials and Methods: SMT data collection below). For chloramphenicol drug treatment in the hupA-PAmCherry strain, additional chloramphenicol was added to a final concentration of 600 µg/mL for 30 minutes before cells were harvested for imaging. For both drug treatments, the drug was added when cells were at mid-log phase (OD600 ~ 0.4). For imaging of the *galP* fluorescent reporter operator system, ACL035 cells were spun down at 4.5 rcf for 5 minutes and resuspended in EZRDM media with 0.4% arabinose to induce expression of the reporter gene. Induction was done at RT for 2 hours with shaking, then cells were again spun down at 4.5 rcf for 5 minutes, washed twice with fresh EZRDM media, then resupended in fresh EZRDM media and grown at 30°C for 1 hour with shaking to allow for maturation of the mCherry fluorophore, after which cells were harvested as described in the section below.

B. Making the HupA-PAmCherry ΔhupB strain

The *HupA-PAmCherry* strain with the *ΔhupB* background (*HupA-PAmCherry ΔhupB*) was used in the Mu phage experiments, to ensure the functionality of the PAmCherry tagged HUα in the absence of HUβ. HupA-PAmCherry strain was transformed with pCP20 plasmid and grown O/N at 42^{o}C to induce kickout of the chloramphenicol resistance cassette (flanked by *frt* sites) along with the temperature sensitive pCP20 plasmid. P1 phage transduction using standard P1vir transduction protocol was performed to transfer *ΔhupB* into the *HupA-PAmCherry* strain, chloramphenicol resistance was used as selection marker.

C. Mu phage assay

Spot dilution plates were performed initially to estimate proper Mu phage lysate dilution to use for phage plaque counting. Cells were grown overnight in Luria Broth (LB) at 37°C, and the next day reinoculated into LB with 1 mM $CaCl_2$ and 2.5 mM $MgCl_2$. Cells were grown until O.D. ~ 0.4, then spun down and resuspended so the final O.D. for later use in the protocol is ~ 1. Next, 100 μl of cells were mixed with 10 μl of the appropriate dilution of Mu phage lysate. This was mixed quickly into 3 ml of 0.7% top agar (warmed to 55°C) and immediately poured onto TB plates (warmed to 37°C) and redistributed evenly prior to hardening. Plates were let dry at RT and incubated overnight at 37°C. The next day, plates were examined and the plaques were counted by eye. The PFU/ml (plaque forming units, Mu/mL) was calculated using the following equation:

$$\frac{PFU}{mL} = \frac{mean(\#\, plaques)}{D \cdot V} \qquad (3.1)$$

where D = dilution and V = volume of diluted virus added to plate.

D. Mini-P1 plasmid transformation assay

Electrocompetent cells were made using standard protocol. In brief, 6 ml of *E. coli* cells were grown until O.D. ~0.4 in LB, the cultures were put on ice for 30 min. Cells were spun down at 4100 rpm at 4°C for 10 min, cell pellet was resuspended in 6 ml ice cold H_2O. Cells were spun down again and resuspended in 3 ml of ice cold 10% glycerol. This was repeated once more, then cells were resuspended in 1.5 ml ice cold 10% glycerol and spun down again. The final cell pellet was resuspended in ~300 *µ*l of GYT media, and 6 individual aliquots of 50 *µ*l of competent cells were made. Cells were quickly frozen in dry ice, and stored at -80°C until later use.

The same batch of competent cells were used for all the transformation experiments for consistency. The pUC19 transformation control was performed in order to normalize the transformation efficiency of competent cells between the different *E. coli* strains. An appropriate volume of transformation outgrowth was plated on appropriate antibiotic selection LB agar plates and incubated O/N at 37°C. The next day, the numbers of colonies on agar plates were counted and recorded. The mini-P1 CFU/*µ*g DNA was normalized to the same strain's pUC19 CFU/*µ*g DNA to account for competent cell transformation efficiency differences.

The WT MG1655 strain was set as the reference standard for comparison with other *E. coli* strains.

E. SMT data collection

Agarose gel pads were made using a low-melting-temperature agarose (SeaPlaque, Lonza) and EZRDM media, with agarose concentration at 3% w/v. 1.5mL of cells were harvested by spinning down in a bench-top microcentrifuge at 4.5 rcf for 5 minutes and resuspended in ~100μL of fresh EZRDM. Cells were pipetted onto agarose pads and sandwiched between the agarose pad and a #1 coverslip as previously described[68]. Immobilized cells were imaged on an Olympus IX71 inverted microscope with a 100x oil objective (NA =1.45) with 1.6x additional magnification. Photons from cells were collected with an Andor EMCCD camera using MetaMorph imaging software (Molecular Devices). Fluorescence from cells was obtained using solid-state lasers at 405nm and 568nm wavelengths (Coherent). All SMT images were collected using 5ms exposure with 1.74ms of cycle time for a total frame length of 6.74ms. Three movies consisting of 2500 frames each were taken for each cell. No cell was imaged longer than five minutes to avoid phototoxicity effects. Activation of the fluorescent proteins was continuous throughout imaging, and no changes were made to either the activation or excitation power throughout the imaging session.

F. SMT analysis

For single molecule tracking analysis, tiff stacks of cell images were imported into the single molecule tracking software UTrack version 3.1[69] within the Matlab 2017a software. We performed the detection and tracking of molecules on individual movies using the Gaussian Mixture Model setting within the UTrack software. For detection, an alpha value of 0.01 was used; for frame-to-frame linking, 15 frames was the maximum time and 0 to 2 pixels (or up to 6 pixels with increased linking cost) was the maximum distance to link trajectories together. Linked trajectories were then filtered by their intensity (between 0.5-1.5x the average single molecule intensity) and their localization within cells (i.e. any trajectory outside of cells was excluded) and exported as MATLAB files. Trajectories were analyzed using custom in-house code (available upon request) to assess their diffusive properties. Exact equations used to analyze the trajectories are explained in more detail in the supplemental information. To build the HMM, we installed and used the software vbSPT, version 1.1 on MATLAB version 2014a[50]. Cells were rotated such that the long cell-axis corresponded to the x-axis. We used only consecutive frame trajectories and only displacements in the x-direction for analysis.

G. Localizations of HU-PAmCherry

Detections of HU-PAmCherry were determined as above using the UTrack software. Cells were rotated such that the long cell axis corresponded to the x-axis, and the middle of the cell corresponded to the origin. We determined the spatial coordinates based on this axis, and normalized the coordinates to a 4 x 1.5 µm standard cell. Because cells are symmetrical about their long and short axes, we took the absolute value of all coordinates to determine the localization in one cell quadrant. We binned these displacements into 100 x 100 nanometer bins. We duplicated this quadrant along both the short and long axis for easier viewing. For comparisons, we first normalized each two dimensional histogram where maximum value was equal to 1, then subtracted the WT normalized histogram from the normalized histogram of each mutant or drug treatment.

H. Total cellular RNA extraction

Cell growth conditions are same as described above. Starting with an identical number of cells for WT condition (no drug treatment), rifampicin treatment (200 μg/mL, 60 min) and Chloramphenicol treatment (600 μg/mL, 60 min). Take volume of cells to an O.D. equivalent of 8.5, pellet cells (10 min, 4100 rpm, 4^{o}C), resuspend the pellet in 1 mL 1xPBS, and transfer to an Eppendorf tube. The cells were pelleted again (2 min, 14000 rpm, 4^{o}C), the pellet was quickly frozen on dry ice and kept at -80°C until the later steps. For RNA extraction, 600 μL of Trizol was added to the cell pellet after thawing, then the protocol from the Direct-zol RNA Miniprep Kit (Zymo Research) was followed, an in-column DNase I digestion was performed as suggested for 15 min. Finally, RNA was eluted with 50 μl of DNase/RNase free H_2O and quantified by measuring on Nanodrop.

I. RNA FISH

Cells were grown the same as for imaging (described above; grown in rich defined media at 24°C). At mid-log phase growth corresponding to OD600 ~ 0.4, cells were fixed in 2.6% paraformaldehyde, 0.8% glutaraldehyde in 1 x PBS for

15 minutes at room temperature. Cells were then washed twice in 1 x PBS and resuspended into a small volume of GTE (2mL of cells were spun into 50uL final volume). Coverslips were treated with 0.1% poly-L-lysine for 5 to 10 minutes, then rinsed and dried. Cells were spotted onto treated coverslips and left to dry for 20 minutes. Non-adhered cells were washed away with PBS. Cells were then treated with ice-cold 70% ethanol and placed at -20°C for 10 minutes to permeabilize the cells. After permeabilization, cells were washed twice in 2 x SSCT (2 x SSC + 0.1% Tween-20) for five minutes each wash. Cells were then prehybridized in pre-hybridization solution (2 x SSC, 40% formamide, 0.1% Tween-20) for 30 to 45 minutes at 37°C. Pre-hybridization solution was removed, then a small volume of hybridization solution (2 x SSC, 40% formamide, dextran sulfate, yeast tRNAs) with 1µM of Cy3-labelled probe was spotted onto the coverslip, then covered with a clean piece of saran wrap to set. Coverslips were submerged in a water bath set to 94°C for two minutes, then immediately placed at 42°C for 15 to 18 hours to hybridize. The probe sequence used was: Cy3-GTTGCCGGATGCGGCGTAAACGCCTTATCCGGCC. Cells were washed in 40% wash solution (2 x SSC, 40% formamide, 0.1% Tween-20) for 30 minutes at 37°C, then washed in 20% wash solution (2 x SSC, 20% formamide, 0.1% Tween-20) for 10 minutes at room temperature, then washed in 2 x SSCT solution for 10 minutes at room temperature, and finally transferred to 2 x SSC solution until imaging. Coverslips were mounted onto imaging chambers and anti-fading media was perfused into the system. Cells were imaged on an Olympus

IX71 inverted microscope with a 100x oil objective (NA =1.45) with 1.6x additional magnification. Photons from cells were collected with an Andor EMCCD camera using MetaMorph imaging software (Molecular Devices). Fluorescence from cells was obtained using solid-state lasers at 568nm wavelength (Coherent). Z-stacks of cells were collected using 50ms exposure per frame with an EM gain of 300. Z-stacks were imported to ImageJ, where cells were manually identified. Fluorescence in each cell was corrected by subtraction of the background (a random area devoid of cells for each image).

V. Supplemental Information

A. Calculation of diffusion coefficients

Several methods were employed to calculate the best number of diffusion coefficients represented in the data and their diffusion coefficients. First, we calculated all single frame displacements along the x-axis. To achieve this, we manually rotated all cells such that the long axis of the cell was defined as the x-axis and rotated all the trajectories by the same angle. We chose to take displacements along the x-axis as there is less likelihood of these particles encountering the barrier of the cell wall [114]. We found the single-frame displacement along the x-axis and plotted the histogram of their distribution binned into a number of bins equal to the square root of total number of displacements. This distribution resembles a Gaussian centered around zero; due to the nature of Brownian motion, displacements on average will be zero and the width of the distribution is related to the number of diffusion coefficients represented in the data and their values. We fit this distribution of single-frame displacements to the following equation:

$$p(r) = \sum_{i=1}^{n} \frac{p_i}{\sqrt{4\pi D_i t}} e^{-r^2/_{4D_i t}} \tag{3.2}$$

Where *r* is the single frame displacement, *n* is the total number of diffusive states represented in the data, D_i is the diffusion coefficient of the *i*th diffusive state, and p_i is the fraction of molecules in the *i*th state. To determine the best number of states, we fit our data to models for one, two, and three diffusive states. We saw no improvement of the fit and dispersion of residuals when we moved from a two-state to a three-state model, so we deemed the two-state model to be the best fit.

Secondly, we calculated the cumulative displacement probability distribution. Similar to above, we calculated the displacement between consecutive frames for each trajectory. To estimate the cumulative displacement probability distribution, we divided the displacements into a number of bins equal to the squre root of the total number of displacements. For each displacement bin, we calculated the fraction of displacements less than or equal to the bin value. This cumulative displacement probability distribution can be fit to the following equation:

$$cdf(r) = 1 - \sum_{i=1}^{n} p_i e^{-x^2/_{4D_i t}} \tag{3.3}$$

where *r* is the single frame displacement, *n* is the total number of diffusive states represented in the data, D*ᵢ* is the diffusion coefficient of the *i*th diffusive state, and

p_i is the fraction of molecules in the *i*th state such that the sum of all pi values equals one. To determine the best number of states, we fit our data to one, two, and three-state models. We saw no improvement of the fit and dispersion of residuals when we moved from a two-state to a three-state model, so we deemed the two-state model to be the best fit.

Lastly, we calculated the apparent diffusion coefficient from each individual displacement. We binned the apparent diffusion coefficients into twelve bins and plotted their distribution. This distribution of apparent diffusion coefficients can be fit to the equation:

$$p(D) = \sum_{i=1}^{n} \frac{a_i}{D_i} e^{-1/D_i} \tag{3.4}$$

where *n* is the number of diffusive states represented in the data, a_i is the fraction of molecules in the *i*th diffusive state, and D_i is the diffusion coefficient of the ith diffusive state. Again, we determined the best number of state by fitting our data to one, two, and three-state models. We saw no improvement of the fit when we moved from a two-state to a three-state model, so again we deemed the two-state model to be the best fit.

B. Calculation of confinement zone from MSD data

First, we calculated the mean-squared displacement (MSD) for 15 time lags (t) by using the following equation:

$$MSD(t) = \frac{1}{N} \sum_{n=1}^{N} (x_n(t) - x_n(0))^2 \tag{3.5}$$

where N is the total number of displacements to be averaged for that timelag, $x_n(0)$ is the initial coordinates of each trajectory, and $x_n(t)$ is the coordinates at time t. For pure Brownian motion, this normally is linear such that MSD(t) = 4Dt; i.e. the slope of the line is proportional to the diffusion coefficient of the molecule. MSD can be thought of as the area that a collection of particles covers in time. Thus, if particles encounter a barrier, then as time approaches infinity, the MSD remains unchanged. Kusumi, et. al. derived the equation for diffusion within a boundary of length L[115], which is:

$$MSD(t) = \frac{L^2}{6} + \frac{16\,L^2}{\pi^4}\sum_{i=1}^{\infty} e^{\frac{-(i\pi\sigma)^2}{L^2 t}} \qquad (3.6)$$

where L is the total width of the boundary, and σ is the average displacement such that $D = \sigma^2/2$. We approximated the sum up to one hundred steps, ensuring we reached convergence, and used non-linear least squares fitting in MATLAB to find the parameters that gave the least squared residuals. To estimate the error in the fitting parameters, we performed a pseudo-bootstrapping method in which we removed a random trajectory from our data set and estimated the parameters again. We repeated this pseudo-bootstrapping one hundred times for each curve; the error reported was the standard deviation of the values from these bootstrapping calculations.

C. Calculation of the rate constants and dissociation constant for HU

Our HMM of HU gives transition probabilities that can be converted into kinetic rate constants. Given a transition matrix A, we can convert these transition probabilities into kinetic rate constants through this relation:

$$A = \begin{bmatrix} a_{ij} & \cdots & ain \\ \vdots & \ddots & \vdots \\ anj & \cdots & ann \end{bmatrix} = exp\left(\begin{bmatrix} kij & \cdots & kin \\ \vdots & \ddots & \vdots \\ knj & \cdots & knn \end{bmatrix} \cdot \Delta t\right) \tag{3.7}$$

where a_{ij} is the transition probability of a molecule transitioning from state i to state j within the frame rate Δt, and k_{ij} is the corresponding kinetic rate between the ith and jth diffusive state. We assigned state I HU-PAmCherry molecules to be bound to the DNA, and state II HU-PAmCherry molecules to be freely diffusing, we can assume that HU binds to the DNA using this reaction:

$$HU + DNA \underset{k_{on}}{\overset{k_{off}}{\longleftarrow\!\!\!\!\longrightarrow}} HU \cdot DNA$$

Assuming that [DNA] >> [HU], we can use a first order approximation, where $K_d = k_{off}[DNA\}/k_{on}$. We assume that each basepair of the DNA provides a unique binding site for the DNA and assume that on average, there are two full chromosomes per *E. coli* cell under our fast growth conditions. Converting the total number of basepairs into molar concentration gives a concentration of DNA around 4mM. Because we know the k_{on} and k_{off} values from our HMM, we can then estimate the dissociation rate, which we determine to be approximately 5mM.

Chapter 4: Conclusions

I. Transcription kinetics *in vivo* match the *in vitro* biochemical values

In Chapter 2, I explore the dynamics of transcription of RNAP using the FRAP methodology. For RNAP-GFP in rich growth media, we found the fluorescence recovery was complex and not well described by a simple two-state model. We reasoned that RNAP-GFP would undergo the typical transcription cycle from freely diffusing to promoter-bound, transitioning into an elongation complex, then dissociating at termination to become freely diffusing once more; this model is based off the model done by Darzacq and colleagues to model RNA Polymerase II transcription using FRAP[116]. We found this model described the recovery very well, and gave promoter escape rate and termination rate that closely matched the values found in the literature[6,58]. To ensure we were truly modeling transcription, we treated our RNAP-GFP cells with rifampicin, a transcription elongation inhibitor, and monitored the FRAP recovery. Indeed, we saw a dramatically different fluorescence recovery profile that was well described by a simple two-state model with slow exchange on and off of the DNA, consistent with the mechanism of action of rifampicin, which traps RNAP on promoter sequences. Interestingly, RNAP has been found to form large clusters in fast growth media. That the FRAP recovery is well described by a simple transcription initiation-elongation-termination cycle implies that these clusters are not stalled RNAP complexes, but rather are actively engaged in transcription. This result matches the observations from Weng, Bohrer, et. al. demonstrating that cluster formation is dependent on transcription activity; under rifampicin treatment, clusters were severely diminished as compared to all other conditions, including rich growth and AsiA-overexpressed conditions[117]. Together, this implies that RNAP clusters are an important, functional aspect of transcription and are not aggregates, stalled complexes, or other transcriptional states.

II. Assigning functional states to diffusive states for DNA-binding proteins

In this thesis, I relied heavily on single molecule tracking as a technique to study the dynamics of two DNA-binding proteins, RNA polymerase and HU. Using many types of traditional analyses, such as plotting the single frame displacements or the individual molecule apparent diffusion coefficient, it is possible to find the number, diffusion coefficients, and occupation percentages of diffusive states present in the data[48]. This information is valuable, but ultimately meaningless unless the researcher can assign a functional state of the molecule to each diffusive state. To achieve this, one can use genetic or drug manipulations to abolish hypothesized states, or use their hypothesis to make predictions.

For example, in both Chapter 2 and 3, I hypothesized the slowest diffusive state of RNAP and HU with diffusion coefficient of 0.1 $\mu m^2/s$ corresponded to a DNA-bound state. To test this hypothesis, we can compare the diffusive behavior of this state to the diffusive behavior of DNA by monitoring an individual DNA

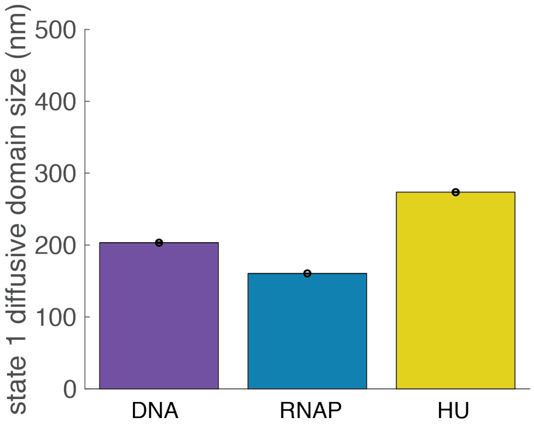


Figure 4.1: Comparison of State 1 diffusive domain sizes for DNA, RNAP-PAmCherry, and HUα-PAmCherry strains.

Diffusive domain sizes for the *tetO* array, RNAP-PAmCherry, and HUα-PAmCherry strains. Values are μ ± 95% confidence interval.

locus via a fluorescence reporter operator system. In our case, we inserted six *tetO* operator sites upstream of the *galP* gene locus and monitored the dynamics of this DNA locus by expressing a mCherry-fused TetR protein from a plasmid. We found that TetR-mCherry had a single diffusive state with diffusion coefficient of ~ 0.09 $\mu m^2/s$, essentially identical to the diffusion coefficient of state 1 for both RNAP and HU. Additionally, when we plotted the MSD of state 1 RNAP and HU, we found the MSD deviated from the typical linear relationship of normal Brownian motion, but instead approached an asymptote typical of Brownian motion under confinement. Using a modified diffusion equation assuming confinement, we determined the diffusive domain size of the DNA locus to be ~ 200nm, which is similar to the values for State 1 RNAP and HU (~160nm and ~230nm respectively, Figure 4.1). These combined pieces of evidence suggest that State 1 RNAP and HU are likely DNA-bound.

The state 2 diffusion coefficient is ~1 order of magnitude less than predicted by the Stokes-Einstein equation at ~ 0.3 – 0.4 µm2/s. We hypothesized that this slowing of the diffusion coefficient was the result of RNAP and HU having weak interactions with the nucleoid. Consistent with this, state 2 RNAP-PAmCherry and HUα-PAmCherry experienced confinement several fold smaller than the area of the nucleoid, with both molecules having ~ 300 nm diffusive domain sizes. We predicted that this confinement was due trapping by the DNA with the size of the diffusive domain corresponding to the pore size of the DNA “meshwork”. To test this hypothesis, we treated cells with chloramphenicol and rifampicin, which condenses and expands the nucleoid respectively. If our hypothesis is true, compaction should decrease the diffusive domain size while expansion should have the opposite effect.

Indeed, we see that for RNAP-PAmCherry, the State 2 diffusive domain size decreases under chloramphenicol treatment and increases under rifampicin treatment. Interestingly, we found that the diffusive domain size remained similar for HU-PAmCherry state 2 under chloramphenicol or rifampicin treatment. We hypothesize that our prior supposition is true, i.e. that state 2 indeed does represent diffusion within the dense nucleoid matrix, but that the size discrepancy between RNAP-PAmCherry and HUα-PAmCherry results in the condensation and expansion having an effect on the much larger RNAP-PAmCherry, but having no effect on the much smaller HUα-PAmCherry (Figure 4.2). This type of mechanism is similar to diffusion of proteins through a gel- if proteins are smaller

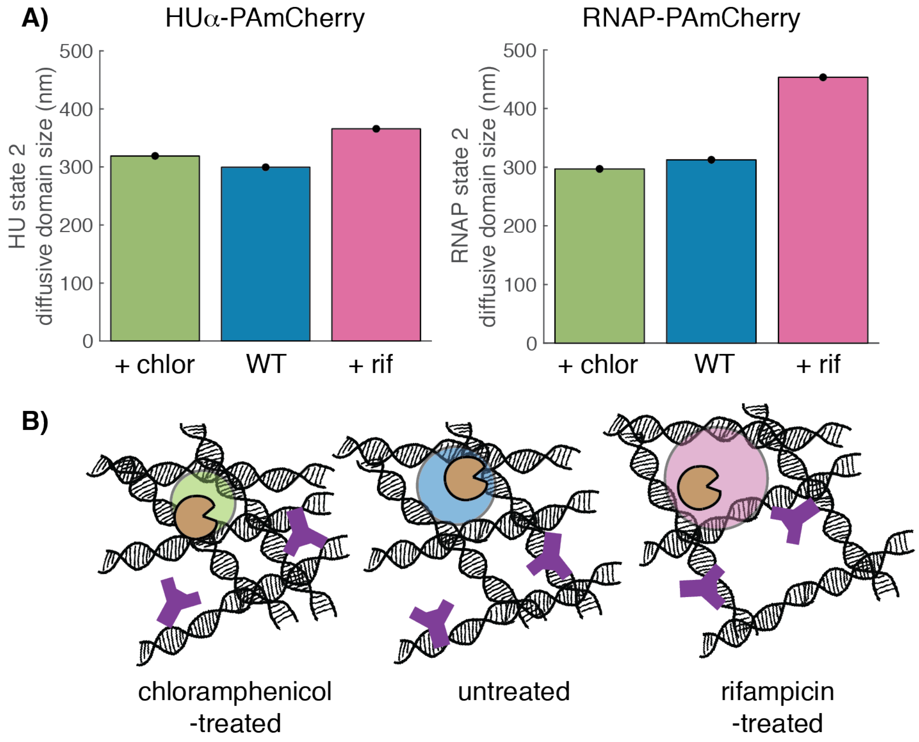


Figure 4.2: State 2 likely represents diffusion within the nucleoid meshwork.

(A) Diffusive domain sizes of state 2 HUα-PAmCherry and state 2 RNAP-PAmCherry in chloramphenicol-treated (green), untreated (blue), and rifampicin-treated conditions. (B) Proposed model of state 2 diffusion: the nucleoid forms a DNA pore meshwork which RNAP-PAmCherry and HUα-PAmCherry can diffuse within. HUα-PAmCherry diffuses similarly in all three conditions due to its small size, while RNAP-PAmCherry feels the effects of nucleoid contraction and expansion.

than the average pore size, their motion is unhindered, but if proteins are similar or larger in size than the average pore size, their motion is more restricted. These results demonstrate the importance of testing the function of each diffusive state for each protein independently.

Last, but not least, for RNAP-PAmCherry in EZRDM, we found a third diffusive state with a much larger diffusion coefficient than States 1 and 2 at ~ 1.4 $\mu m^2/s$. Additionally its diffusive domain size was much larger, at ~ 500nm. We attributed this to RNAP-PAmCherry undergoing free diffusion. Corroborating this, a study done by Stracy and colleagues inhibited DNA replication, resulting in cell elongation containing DNA-free endcaps[60]. In this region, the diffusion coefficient distribution was broad with an average around ~ 2.0 $\mu m^2/s$, similar to our reported value, corroborating our state assignment.

III. Kinetics show mostly non-steady state dynamics

A powerful aspect of SMT is the ability to extract transition kinetics between diffusive states. To avoid the bias of using threshold-based methods, we turned to statistical modeling to form a hidden Markov model (HMM) of our data using the single-frame displacements of all trajectories. For RNAP-PAmCherry, we found the transition rates funneled RNAP-PAmCherry towards the DNA-bound state, as demonstrated by the dwell times of each state- 200ms, 80ms, and 30ms for states 1, 2, and 3, respectively. Numerically solving the rate equations based on the kinetic rates from the HMM gives occupation percentages that differ from the values independently determined by the HMM, indicating that RNAP-PAmCherry in rich growth media may not be steady state. Additionally, HUα-PAmCherry also had HMM-determined occupation percentages that differed from what would be predicted by the independently-determined kinetic rates. State 1 HUα-PAmCherry represented 45% of all HU while State 2 HUα-PAmCherry represented 55% of all HU. However, the kinetic rates determined by the HMM would predict 57% of HU in State 1, and 43% of HU in State 2. This type of discrepancy between the HMM-determined kinetic rates and the independently determined occupation percentages was a common theme for all other conditions imaged for both RNAP-PAmCherry and HUα-PAmCherry.

We predict this discrepancy is likely due to effects of crowding. Theoretical studies demonstrate that under crowding conditions, the association and dissociation rate on and off DNA follows a power law instead of an exponential law as expected from mass action[44,118]. These long tailed distributions will skew the estimations of rate constants in a manner dependent on the power law exponent. Additionally, power law distributions of dissociation times have been observed previously for TetR in U2OS cells[43]. If true, this would be interesting given that we see diffusive behavior consistent with normal Brownian motion within a confined environment, and not anomalous diffusion. However, because our trajectories are short, and thus we cannot fit our MSD to a large number of timelags, it is possible that by imaging at longer timepoints, we could see the diffusive behavior be described by anomalous diffusion within a confined environment, or change behavior at different timescales as previously seen for chromosomal loci dynamics[33]. Interestingly, this trend of occupation percentages that disagreed with the state-switching kinetics occurred for RNAP and HU under a wide variety of conditions, suggesting this is a feature of their surroundings and not necessarily inherent to the protein itself. Despite the attractiveness of such a model, the underlying cause of the non-steady state dynamics remains elusive and needs follow up work.

IV. Metabolism alters the transcription cycle but not the search dynamics

In EZRDM media, transcription is mostly dedicated to ribosomal operons. These operons have a strong promoter and are predicted to have more RNAP per gene than an average gene. To test whether a different metabolic state in which *rrn* synthesis was abolished or reduced would alter the transcription kinetics, we performed FRAP and SMT on tagged RNAP grown in M9 media, or with the T4 phage anti-σ70 factor AsiA overexpressed. Both M9 and AsiA overexpression conditions displayed dynamics that were best described by the three state model described in Chapter 4, Section I (above). In these conditions, the dynamics were altered such that the freely diffusing population of RNAP increased. For M9, the escape rate and termination rates increased, leading to a similar percentage of transcribing RNAP as the rich growth condition. For the AsiA overexpression condition, the escape rate decreased while the termination rate increased, leading to a lower percentage of transcribing RNAP. These results were consistent with the expectations based on the literature- as RNAP is shifted towards mRNA synthesis in both M9 and AsiA overexpression conditions, we expect the relative number of freely diffusing RNAP to increase. These data suggest that the metabolic rate does influence transcription kinetics.

Interestingly, while the kinetics of the transcription cycle were affected by the metabolic state, the effects on the search process were much more subtle. While there was no cytoplasmically diffusing RNAP in these two conditions, the diffusion coefficients of the first two states were essentially identical to the diffusion coefficients of the first two states in rich growth media. Additionally, the dwell times of RNAP in states 1 and 2 for all three conditions were comparable.

V. Using SMT to uncover the molecular mechanisms of HU-mediated chromosome organization

A. Dynamics of WT HUα-PAmCherry show transient dynamics

In Chapter 3, I utilized the SMT methodology to understand the molecular mechanisms of HU-mediated chromosome organization. HU has two subunits, α and β, that can form either HUαα homodimers of HUαβ heterodimers; HUββ homodimers are not detectible to a large extent. It has long been proposed that these two types of HU dimers have differential effects on chromosome organization. Additionally, it was recently demonstrated that HU acts as a chaperone to deposit small non-coding RNAs onto cruciform DNA structures. Puzzlingly, HU has poor affinity for dsDNA but high affinity for recombination intermediates such as knicked, gapped, or stemloop nucleic acids. Additionally, early crystal structures of HU demonstrate that it either binds to bent DNA or bends DNA when it binds. Crystal structures of non-specifically bound HU show that HU can form a dimer of dimers, enabling these dimer of dimers to bundle DNA together. The authors of this study propose that multimerization of these dimer of dimers can facilitate chromosomal organization by creating a HU dimer-DNA network[90]. These disparate observations have yet failed to coalesce into a working model for HU-mediated chromosome organization.

To probe into the molecular mechanisms of HU, we tracked the positions of HUα-PAmCherry with high temporal speed. As mentioned in Section II, HU displayed diffusive states similar to other DNA-binding proteins, such as RNAP in our study and LacI[41] or UvrA[64] in other studies. Consistent with prior biochemical studies, dynamics of HUα-PAmCherry were transient, with dwell times of ~100ms in state 1 and ~75ms in state 2, leading to an estimation of the K_d to be ~ 5 mM. While consistent with biochemical assays on binding to linear dsDNA, this is inconsistent with the multimerization model, which would suggest HU multimers form stable, long-lived multimer-DNA networks. It is possible these multimerization networks exist *in vivo*, but undergo constant exchange. However, our PALM data of HUα-PAmChery saw nearly homogenous distribution of HUα-PAmCherry inconsistent with multimer-DNA network formation, suggesting either these structures are relatively small, or do not contribute significantly to chromosomal organization. Several studies have demonstrated the critical factor HU plays in chromosomal organization, thus the mechanism of HU-mediated chromosomal organization may be the collective mass action of thousands of transient interactions. In nature, army ants can swarm and kill small vertebrate animals through the collective mass action of thousands of swarming ants. This is analogous to the proposed mechanism of HU-mediated chromosomal organization, where thousands of HU “swarm” the DNA and their collective action shapes the *E. coli* chromosome.

B. HUαα displays differential dynamics to HUαβ

To gain further insight, we tracked HUα-PAmCherry under several genetic perturbations. We found that HUαα homodimers, by tracking HUα-PAmCherry in a ΔhupB strain background, exhibited differential dynamics to the WT condition, with overall increased diffusion coefficients. The dynamics remained transient, and while HUαα still localized to the nucleoid, it localized to the polar edge of the nucleoid while WT HUα localized more towards the midcell. These results suggested that the theory in the field that the two HU dimers have differential roles in chromosome organization might be true, as the two subunits demonstrate significantly different dynamics and localizations.

C. HUαα is likely responsible for non-specific DNA binding

Abolishing non-specific binding by mutating surface lysines to alanines, we saw a dramatic change in HU dynamics: only 5% of all HU were able to bind to DNA, while the remaining 95% of HU not only had a significantly increased diffusion coefficient compared to WT cells but also exhibited localizations that emcompassed the entire cell rather than strict nucleoid localization. Interestingly, when we examined this triple lysine mutant in a ΔhupB background, we observed that while the diffusion coefficients increased, the triple lysine mutant HUα regained partial function to bind to the nucleoid. This suggested that when this triple lysine mutant HUα is bound to HUβ, it renders the heterodimer nonfunctional. The triple lysine HUαα homodimer still displays dramatic changes- in fact, the growth rate and cell length were longer than deletion of both HU subunits- but without HUβ, can regain partial function. This (and the data below) suggests that HUαα is primarily responsible for binding non-specifically to DNA.

D. HUαβ is likely responsible for structure-specific binding

Abolishing structure-specific binding by mutating the universally conserved proline responsible for intercalating and stabilizing bent DNA, we observed a more moderate change in dynamics. Strikingly, the dynamics of the P63A mutant HUα highly resembled that of the WT HUαα homodimers- i.e. their diffusion coefficients increased and their localizations shifted from more midcell localized to more polar localized. Even more surprisingly, deletion of the HUβ subunit saw only modest effects on HU dynamics and localization compared to both WT HUαα homodimers and the P63A mutant HUα in the presence of HUβ. This would suggest that the P63A mutant HUα cannot form HUαβ heterodimers. Moreover, the HUα(P63A) mutant strain with HUβ present displayed relatively normal growth rates and cell lengths, but without HUβ grew slower and had dramatically longer cell lengths. However, the cell length was able to recover when the cells were grown at 37°C, suggesting that the structure-specific binding is critical for cold-shock gene regulation. Additionally, it suggests that HUβ is necessary for this structure-specific gene regulation, implying that HUαβ is primarily responsible for structure-specific DNA binding.

E. HU-naRNA interaction plays a still elusive role in chromosome organization

Lastly, deletion of the non-coding RNA binding partner from the REP element *REP325* demonstrated a modest effect on HU dynamics. The diffusion coefficients did not change, but the relative percentage of HU in the DNA-bound state decreased. However, the localizations of HU in the *ΔREP325* background shifted from midcell localized to more quarter-cell localized, with expansions along the y-axis compared to WT. Previously it was reported that *ΔREP325* exhibits an expanded nucleoid. Thus, it is possible the changes in dynamics we observed were merely the result of nucleoid reorganization. However, HUα-PAmCherry under chloramphenicol treatment, where the nucleoid is compacted, displayed essentially identical dynamics to the untreated HUα-PAmCherry. Rifampicin-treated HUα-PAmCherry, where the nucleoid is expanded, displayed dynamics almost identical to that of HUα-PAmCherry in the *ΔREP325* background, suggesting the change in dynamics was due to the loss of the naRNAs as opposed to the nucleoid structure. This was surprising, given that when the *REP325* locus was deleted, it was previously observed that some contacts between cruciform DNA structures were eliminated[87]. The modest effects of *ΔREP325* on HUα-PAmCherry dynamics suggest that this role of HU may be less important. However, cruciform DNA structures are ubiquitous in all trees of life[112], suggesting that this role remains poorly understood.

F. SMT highlights importance of cellular context

This study highlights the powerful nature of SMT to elucidate molecular mechanisms; SMT contains a plethora of rich information, from diffusion coefficients to kinetics to localizations coded by diffusive state. Such information allows for a detailed decomposition of the dynamics of proteins under various conditions, such as HU under various genetic manipulations, allowing us to provide unparalleled insight into a decades-old problem. The vast majority of prior literature on HU-mediated chromosomal organization was performed *in vitro* and as such, may not be relevant within a heterogeneous, crowded environment such as the cell cytoplasm and/or nucleoid. The fact that our observations demonstrate that HU is in fact relatively homogenously distributed, and has transient interactions with DNA suggests that static models wherein HU forms HU-DNA networks are likely incorrect within an *in vivo* context. Moreover, by monitoring HU under differing genetic mutations, we were able to elucidate that non-specific binding is its most important role, followed by structure-specific binding for gene regulation, and finally that naRNA deposition plays an as of yet unidentified role.

VI. Future Directions

A. RNAP transcription

Using FRAP, we were able to model the recovery of fluorescence based on a three-state model depicting the typical transcription initiation-elongation-termination cycle. To minimize the effects of the heterogeneity of the distribution of RNAP, we averaged many cells together for the model. However, analysis techniques exist to monitor the recovery of individual cells. In our case, photobleaching is a huge concern, leading to systematically larger errors we monitor the fluorescence recovery. By using better fluorophores, such as HaloTag and Janelia modified rhodamine dyes, we should be able to monitor the fluorescence recovery with significantly higher accuracy throughout the entire ten minute data collection period. With this improvement in resolution, we can use models that take into account the initial steady state distribution as well as diffusion to model the recovery of individual cells[56].

Additionally, our three-state model combines promoter-bound RNAP with non-specifically bound RNAP, meaning the on and off rates from the DNA are a combined rate of these two processes. The limiting factor in our study was computational power- with a forward finite difference method as we used, a large number of steps is required to obtain accurate numerical integration results for the FRAP simulations. Adding in a fourth state greatly increases the number of timesteps required which becomes computationally prohibitive on a standard computer. There are two ways to get around this issue: (1) employ the backwards finite difference which requires substantially fewer timesteps but is more computationally intensive per timestep, and (2) use a computational core to parallelize modeling of multiple individual cells. Another possible way to speed the process is to switch to C++ or python programming languages, which often offer significantly faster computational speeds than MATLAB. Having models for tens of individual cells should show the distribution in kinetic rates for the transcription cycle. These distributions could lead to valuable insight into transcriptional noise or general gene expression kinetics.

B. RNAP search process

Our studies found that RNAP experienced significant confinement within the cell, from DNA-bound, diffusing within the nucleoid, to diffusing within the cytoplasm. Because other proteins in *E. coli* and *C. crescentus* experience confinement[119], I suspect the confinement observed for both RNAP and HU are due to some property of the cytoplasm and/or of DNA-binding proteins in general. To further probe the physical nature of the cytoplasm, it would be fruitful to compare the diffusion of DNA-binding proteins to other types of proteins; the relative similarity or difference between these diffusive behaviors would elucidate the relative strengths of the general weak interactions of the crowded cytoplasm over the interactions imparted by the function of the protein.

Puzzlingly, we found that both RNAP and HU deviated from steady state dynamics in every imaged condition, from slow growth to fast growth and under various drug perturbations. This deviation from steady state dynamics could be due to a multitude of factors, but because multitudes of theoretical studies and some live cell studies have shown that crowding can lead to power law distributions in dwell times in various states, I suspect the deviation from steady state is due to this crowding effect. To probe this hypothesis, several things can be done. First, one can use blurred motion single particle tracking to observe only bound or slowly moving populations of RNAP and HU. By utilizing a better fluorophore such as the Janelia modified rhodamine dyes bound to HaloTag, or by electroporating in dye-labelled protein, it is possible to monitor the true survival time distribution of a protein and discern the shape of its distribution. Second, one can calculate the mean first passage time of molecules across the entire cell; for well-mixed systems and Brownian diffusers, the mean first passage time follows the typical expected time from the diffusion equation, but shows a steeper length dependence when molecules undergo anomalous diffusion[42]. Third, one can monitor the distribution of angles between consecutive timelag displacements. For normal Brownian diffusers, the distribution of angles between displacements is homogenous. However, for molecules undergoing confinement or anomalous diffusion, there will be a preference towards backwards angles. By monitoring the angles over several timelags, it is possible to better elucidate the diffusive behavior of RNAP[42]. Monitoring displacement angles requires millions of displacements, meaning that a higher throughput set-up for collecting trajectories and/or using HaloTag in combination with the Janelia Farms modified rhodamine dyes to increase the trajectory length will be required. Together, these techniques should provide even more detailed information into the diffusive behavior of RNAP.

Finally, the dynamics of searching RNAP can be monitored as a function of crowding by overexpressing a non-endogenous protein to varying degrees. If crowding is an important factor, one would expect the search dynamics to change in a manner dependent on the average concentration of the non-endogenous protein. By utilizing SMT and the methods mentioned above, a clear trend should be able to be discerned.

C. HU-mediated chromosome organization

Using SMT, we were able to interrogate the role of various DNA-binding modes of HU on the dynamics and localization of HU. We found that HUαα homodimers and HUαβ heterodimers displayed differential dynamics.

Additionally, using genetic manipulations we were able to discern that HUαα is likely more involved in non-specific DNA binding, while HUαβ is likely more involved in structure-specific binding. In the 2D histogram of localizations, HUαβ appeared to localize more towards the mid-cell while Huαα localized more towards the polar edge of the nucleoid. Performing two-color PALM through fusion of two different fluorescent proteins to HUα and HUβ respectively could corroborate these observations.

Surprisingly, despite the distinctive phenotype observed when *REP325* was deleted, we found the dynamics of HU experienced only a modest change in a *ΔREP325* background. Limited information exists on the relative abundance of naRNAs from the *REP325*. We attempted to perform smFISH experiments on the naRNAs produced from this locus but were met with limited success. One issue to overcome is the fact that these naRNAs are highly homologous (up to ~50% homology) to every other REP element in the chromosome. Additionally, each naRNA is only ~ 75bp long, several orders of magnitude smaller than molecules typically targeted for smFISH. One promising avenue that we didn’t have time to follow up on was using FRET FISH, where two probes against the naRNA with fluorophores able to undergo FRET were used to minimize the off-target effects from homology with other sequences. I was able to optimize the binding of one probe but not the other. If someone was able to optimize the hybridization of the second FRET FISH probe, then smFISH could be performed to discern the approximate abundance and localization of naRNAs in cells. Furthermore, these levels could be monitored in the two DNA-binding mutants to see if there is a correlation between the two.

Additionally, because the naRNA is small, it is a perfect candidate for electroporation. By electroporating dye-labelled naRNAs into the cell and monitoring their movement, their function within the cell could be further elucidated. This data in combination with doing a sequencing technique that monitors DNA-RNA contacts, detailed localization and dynamical information on these naRNAs could be determined. Because our SMT of HU in the *ΔREP325* background failed to elucidate the role of naRNAs, this could provide a better angle to probe this function of HU.

Finally, HU displayed transient dynamics on the DNA, suggesting that HU is not stably bound to DNA. Furthermore, PALM images demonstrated surprisingly homogenous distributions of HU within the nucleoid, suggesting no clustering of HU was present. We proposed that the collective action of thousands of transient interactions on the DNA was responsible for chromosome organization. Fortunately, the HU-DNA interaction has been studied by optical tweezers[89], thus the properties of DNA as a function of HU concentration are known. Using this data in combination with the SMT in Chapter 3, one could build a polymer model of the chromosome broken into segments that display different physical properties depending on whether HU is bound to that segment or not. Once a model for WT HU was built, optical tweezers experiments could be performed on the various HU mutants and again modeled to corroborate the observations in our SMT studies.

VII. Final Thoughts

There are a few key takeaways from this thesis: (1) Dynamics of proteins in live cells do not always match the dynamics measured in test tubes, (2) SMT is powerful and capable of elucidating complex molecular mechanisms so long as you have proper controls and state assignments, (3) Protein diffusion within live bacterial cells at the very least experiences significant confinement and appears to affect the kinetics of state switching. In the future, researchers endeavoring to perform SMT should take care to perform careful controls in order to correctly assign a functional state to a diffusive state. Additionally, researchers should move away from using photoactivatable fluorescent proteins and move towards using organic dyes, which are infinitely more photostable and provide trajectories that last on the order of seconds to minutes as opposed to hundreds of milliseconds to seconds.

Moreover, SMT is an extremely powerful tool that contains a rich array of information. As such, it often requires careful consideration of the physical properties of proteins, especially their diffusion, and the soft matter properties of the cytoplasm and/or nucleus. However, there exists a large discrepancy between the number of theoretical studies on protein diffusion within a crowded environment and the application of these theories on real experimental data. This points to a large void in collaborations between theoretical physicists and experimental biophysicists. Huge opportunity is available for those willing to bridge these two disciplines.

Another reason I think SMT is powerful comes from a philosophy I have about biology: everything comes back to diffusion. No biochemical reaction can occur without diffusion, and diffusion is often underappreciated. Textbooks often depict protein-protein interactions as two proteins simply coming together as if they were pulled, but the reality is that proteins must perform a blind drunken walk to find each other. This randomness leads to stochastic events in the cell, which leads to gene expression noise and even cell fate determination. Thus, I hope more emphasis is placed on the importance of diffusion in biological processes as the field of biophysics progresses.

References

1. Lee, D. J., Minchin, S. D. & Busby, S. J. W. Activating transcription in bacteria. *Annu. Rev. Microbiol.* **66,** 125–152 (2012).
2. Kaczanowska, M. & Rydén-Aulin, M. Ribosome biogenesis and the translation process in Escherichia coli. *Microbiol. Mol. Biol. Rev.* **71,** 477–494 (2007).
3. deHaseth, P. L., Lohman, T. M., Burgess, R. R. & Record, M. T. Nonspecific interactions of Escherichia coli RNA polymerase with native and denatured DNA: differences in the binding behavior of core and holoenzyme. *Biochemistry* **17,** 1612–1622 (1978).
4. Gourse, R. L. *et al.* Strength and regulation without transcription factors: lessons from bacterial rRNA promoters. *Cold Spring Harb. Symp. Quant. Biol.* **63,** 131–139 (1998).
5. Jishage, M., Iwata, A., Ueda, S. & Ishihama, A. Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of four species of sigma subunit under various growth conditions. *Journal of Bacteriology* **178,** 5447–5451 (1996).
6. Hsu, L. M., Vo, N. V., Kane, C. M. & Chamberlin, M. J. In vitro studies of transcript initiation by Escherichia coli RNA polymerase. 1. RNA chain initiation, abortive initiation, and promoter escape at three bacteriophage promoters. *Biochemistry* **42,** 3777–3786 (2003).
7. Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314,** 1139–1143 (2006).
8. Harden, T. T. *et al.* Bacterial RNA polymerase can retain σ70 throughout transcription. *Proc. Natl. Acad. Sci. U.S.A.* **113,** 602–607 (2016).
9. Stepanova, E. *et al.* Analysis of promoter targets for Escherichia coli transcription elongation factor GreA in vivo and in vitro. *Journal of Bacteriology* **189,** 8772–8785 (2007).
10. Nudler, E. & Gottesman, M. E. Transcription termination and anti-termination in E. coli. *Genes Cells* **7,** 755–768 (2002).
11. Schmidt, M. C. & Chamberlin, M. J. nusA protein of Escherichia coli is an efficient transcription termination factor for certain terminator sites. *J. Mol. Biol.* **195,** 809–818 (1987).
12. Li, J., Mason, S. W. & Greenblatt, J. Elongation factor NusG interacts with termination factor rho to regulate termination and antitermination of transcription. *Genes Dev.* **7,** 161–172 (1993).

13. Burns, C. M., Richardson, L. V. & Richardson, J. P. Combinatorial effects of NusA and NusG on transcription elongation and Rho-dependent termination in Escherichia coli. *J. Mol. Biol.* **278,** 307–316 (1998).
14. Adam, G. & Delbruck, M. Reduction of dimensionality in biological diffusion processes. *Structural chemistry and molecular biology* 198–215 (1963).
15. McCloskey, M. A. & Poo, M. M. Rates of membrane-associated reactions: reduction of dimensionality revisited. *J. Cell Biol.* **102,** 88–96 (1986).
16. Hoł yst, R., Błażejczyk, M., Burdzy, K., Goralski, G. & Bocquet, L. Reduction of dimensionality in a diffusion search process and kinetics of gene expression. *Physica A: Statistical Mechanics and its Applications* **277,** 71–82 (2000).
17. Li, G.-W., Berg, O. G. & Elf, J. Effects of macromolecular crowding and DNA looping on gene regulation kinetics. *Nature Physics* **5,** 294 EP –
18. Hammar, P. *et al.* The lac repressor displays facilitated diffusion in living cells. *Science* **336,** 1595–1598 (2012).
19. Marcovitz, A. & Levy, Y. Obstacles may facilitate and direct DNA search by proteins. *Biophysical Journal* **104,** 2042–2050 (2013).
20. Shvets, A., Kochugaeva, M. & Kolomeisky, A. B. Role of Static and Dynamic Obstacles in the Protein Search for Targets on DNA. *J Phys Chem B* **120,** 5802–5809 (2016).
21. van den Broek, B., Lomholt, M. A., Kalisch, S.-M. J., Metzler, R. & Wuite, G. J. L. How DNA coiling enhances target localization by proteins. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 15738–15742 (2008).
22. Lomholt, M. A., van den Broek, B., Kalisch, S.-M. J., Wuite, G. J. L. & Metzler, R. Facilitated diffusion with DNA coiling. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 8204–8208 (2009).
23. Zhou, H.-X. & Szabo, A. Enhancement of association rates by nonspecific binding to DNA and cell membranes. *Phys. Rev. Lett.* **93,** 178101 (2004).
24. Mahmutovic, A., Berg, O. G. & Elf, J. What matters for lac repressor search in vivo--sliding, hopping, intersegment transfer, crowding on DNA or recognition? **43,** 3454–3464 (2015).
25. Pollak, A. J., Chin, A. T. & Reich, N. O. Distinct facilitated diffusion mechanisms by E. coli Type II restriction endonucleases. *Biochemistry* **53,** 7028–7037 (2014).
26. Bancaud, A. *et al.* Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. *The EMBO Journal* **28,** 3785–3798 (2009).

27. LAURENT, T. C. THE INTERACTION BETWEEN POLYSACCHARIDES AND OTHER MACROMOLECULES. 5. THE SOLUBILITY OF PROTEINS IN THE PRESENCE OF DEXTRAN. *Biochem. J.* **89,** 253–257 (1963).
28. Fuller, R. S., Kaguni, J. M. & Kornberg, A. Enzymatic replication of the origin of the Escherichia coli chromosome. *Proc Natl Acad Sci USA* **78,** 7370–7374 (1981).
29. Shtilerman, M. D., Ding, T. T. & Lansbury, P. T. Molecular crowding accelerates fibrillization of alpha-synuclein: could an increase in the cytoplasmic protein concentration induce Parkinson's disease? *Biochemistry* **41,** 3855–3860 (2002).
30. Batra, J., Xu, K., Qin, S. & Zhou, H.-X. Effect of macromolecular crowding on protein binding stability: modest stabilization and significant biological consequences. *Biophysical Journal* **97,** 906–911 (2009).
31. Elowitz, M. B., Surette, M. G., Wolf, P. E., Stock, J. B. & Leibler, S. Protein mobility in the cytoplasm of Escherichia coli. *Journal of Bacteriology* **181,** 197–203 (1999).
32. Golding, I. & Cox, E. C. Physical nature of bacterial cytoplasm. *Phys. Rev. Lett.* **96,** 098102 (2006).
33. Weber, S. C., Spakowitz, A. J. & Theriot, J. A. Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. *Phys. Rev. Lett.* **104,** 238102 (2010).
34. Minton, A. P. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.* **276,** 10577–10580 (2001).
35. Wenner, J. R. & Bloomfield, V. A. Crowding effects on EcoRV kinetics and binding. *Biophysical Journal* **77,** 3234–3241 (1999).
36. Berry, H. Monte carlo simulations of enzyme reactions in two dimensions: fractal kinetics and spatial segregation. *Biophysical Journal* **83,** 1891–1901 (2002).
37. Grima, R. & Schnell, S. A systematic investigation of the rate laws valid in intracellular environments. *Biophys. Chem.* **124,** 1–10 (2006).
38. Cravens, S. L. *et al.* Molecular crowding enhances facilitated diffusion of two human DNA glycosylases. **43,** 4087–4097 (2015).
39. Ellis, R. J. Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.* **26,** 597–604 (2001).
40. Weiss, M., Elsner, M., Kartberg, F. & Nilsson, T. Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells. *Biophysical Journal* **87,** 3518–3524 (2004).

41. Elf, J., Li, G.-W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316,** 1191–1194 (2007).
42. Izeddin, I. *et al.* Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *eLife Sciences* **3,** 23352 (2014).
43. Normanno, D. *et al.* Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat Commun* **6,** 7357 (2015).
44. Zaid, I. M., Lomholt, M. A. & Metzler, R. How subdiffusion changes the kinetics of binding to a surface. *Biophysical Journal* **97,** 710–721 (2009).
45. Friedman, L. J., Mumm, J. P. & Gelles, J. RNA polymerase approaches its promoter without long-range sliding along DNA. *Proc Natl Acad Sci USA* **110,** 9740–9745 (2013).
46. Wang, F. *et al.* The promoter-search mechanism of Escherichia coli RNA polymerase is dominated by three-dimensional diffusion. *Nature Structural & Molecular Biology* **20,** 174–181 (2013).
47. Sage, D. *et al.* Quantitative evaluation of software packages for single-molecule localization microscopy. *Nat Meth* **12,** 717–724 (2015).
48. Vrljic, M., Nishimura, S. Y. & Moerner, W. E. Single-molecule tracking. *Methods Mol. Biol.* **398,** 193–219 (2007).
49. Das, R., Cairo, C. W. & Coombs, D. A Hidden Markov Model for Single Particle Tracks Quantifies Dynamic Interactions between LFA-1 and the Actin Cytoskeleton. *PLoS Comput Biol* **5,** e1000556–16 (2009).
50. Persson, F., Lindén, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat Meth* **10,** 265–269 (2013).
51. Schulz, J. H. P., Chechkin, A. V. & Metzler, R. Correlated continuous time random walks: combining scale-invariance with long-range memory for spatial and temporal dynamics. *Journal of Physics A Mathematical and Theoretical* **46,** 475001 (2013).
52. Buc, H. & McClure, W. R. Kinetics of open complex formation between Escherichia coli RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps. *Biochemistry* **24,** 2712–2723 (1985).
53. Davis, C. A., Bingman, C. A., Landick, R., Record, M. T. & Saecker, R. M. Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by Escherichia coli RNA polymerase. *Proc Natl Acad Sci USA* **104,** 7833–7838 (2007).

54. Reppas, N. B., Wade, J. T., Church, G. M. & Struhl, K. The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. *Molecular Cell* **24,** 747–757 (2006).
55. Zimmerman, S. B. & Trach, S. O. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J. Mol. Biol.* **222,** 599–620 (1991).
56. Beaudouin, J., Mora-Bermúdez, F., Klee, T., Daigle, N. & Ellenberg, J. Dissecting the Contribution of Diffusion and Interactions to the Mobility of Nuclear Proteins. *Biophysical Journal* **90,** 1878–1894 (2006).
57. Nakamura, Y. & Yura, T. Effects of rifampicin on synthesis and functional activity of DNA-dependent RNA polymerase in Escherichia coli. *Mol. Gen. Genet.* **145,** 227–237 (1976).
58. Gotta, S. L., Miller, O. L. & French, S. L. rRNA transcription rate in Escherichia coli. *Journal of Bacteriology* **173,** 6647–6649 (1991).
59. Cabrera, J. E., Cagliero, C., Quan, S., Squires, C. L. & Jin, D. J. Active transcription of rRNA operons condenses the nucleoid in Escherichia coli: examining the effect of transcription on nucleoid structure in the absence of transertion. *Journal of Bacteriology* **191,** 4180–4185 (2009).
60. Stracy, M. *et al.* Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proc. Natl. Acad. Sci. U.S.A.* **112,** E4390–9 (2015).
61. Liu, L. & Luo, K. Molecular crowding effect on dynamics of DNA-binding proteins search for their targets. *J Chem Phys* **141,** 225102 (2014).
62. Ridgway, D. *et al.* Coarse-Grained Molecular Simulation of Diffusion and Reaction Kinetics in a Crowded Virtual Cytoplasm. *Biophysical Journal* **94,** 3748–3759 (2008).
63. Sharma, U. K. & Chatterji, D. Differential mechanisms of binding of anti-sigma factors Escherichia coli Rsd and bacteriophage T4 AsiA to E. coli RNA polymerase lead to diverse physiological consequences. *Journal of Bacteriology* **190,** 3434–3443 (2008).
64. Stracy, M. *et al.* Single-molecule imaging of UvrA and UvrB recruitment to DNA lesions in living Escherichia coli. *Nat Commun* **7,** 12568 (2016).
65. Ando, T., Yu, I., Feig, M. & Sugita, Y. Thermodynamics of Macromolecular Association in Heterogeneous Crowding Environments: Theoretical and Simulation Studies with a Simplified Model. *J Phys Chem B* **120,** 11856–11865 (2016).
66. Zhou, H.-X. Rapid search for specific sites on DNA through conformational switch of nonspecifically bound proteins. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 8651–8656 (2011).

67. Tabaka, M., Kalwarczyk, T. & Hołyst, R. Quantitative influence of macromolecular crowding on gene regulation kinetics. **42**, 727–738 (2014).
68. Buss, J. *et al.* In vivo organization of the FtsZ-ring by ZapA and ZapB revealed by quantitative super-resolution microscopy. *Mol. Microbiol.* **89**, 1099–1120 (2013).
69. Jaqaman, K. *et al.* Robust single-particle tracking in live-cell time-lapse sequences. *Nat Meth* **5**, 695–702 (2008).
70. Endesfelder, U., Malkusch, S., Fricke, F. & Heilemann, M. A simple method to estimate the average localization precision of a single-molecule localization microscopy experiment. *Histochem. Cell Biol.* **141**, 629–638 (2014).
71. Sinden, R. R. & Pettijohn, D. E. Chromosomes in living Escherichia coli cells are segregated into domains of supercoiling. *Proc Natl Acad Sci USA* **78**, 224–228 (1981).
72. Boccard, F., Esnault, E. & Valens, M. Spatial arrangement and macrodomain organization of bacterial chromosomes. *Mol. Microbiol.* **57**, 9–16 (2005).
73. Browning, D. F., Grainger, D. C. & Busby, S. J. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Current Opinion in Microbiology* **13**, 773–780 (2010).
74. Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S. & Ishihama, A. Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. *Journal of Bacteriology* **181**, 6361–6370 (1999).
75. Claret, L. & Rouviere-Yaniv, J. Variation in HU composition during growth of Escherichia coli: the heterodimer is required for long term survival. *J. Mol. Biol.* **273**, 93–104 (1997).
76. Kamashev, D. & Rouviere-Yaniv, J. The histone-like protein HU binds specifically to DNA recombination and repair intermediates. *The EMBO Journal* **19**, 6527–6535 (2000).
77. Dixon, N. E. & Kornberg, A. Protein HU in the enzymatic replication of the chromosomal origin of Escherichia coli. *Proc Natl Acad Sci USA* **81**, 424–428 (1984).
78. Bonnefoy, E. & Rouviere-Yaniv, J. HU, the major histone-like protein of E. coli, modulates the binding of IHF to oriC. *The EMBO Journal* **11**, 4489–4496 (1992).
79. Prieto, A. I. *et al.* Genomic analysis of DNA binding and gene regulation by homologous nucleoid-associated proteins IHF and HU in Escherichia coli K12. **40**, 3524–3537 (2012).

80. Balandina, A., Claret, L., Hengge-Aronis, R. & Rouviere-Yaniv, J. The Escherichia coli histone-like protein HU regulates rpoS translation. *Mol. Microbiol.* **39,** 1069–1079 (2001).
81. Becker, N. A., Kahn, J. D. & Maher, L. J. Effects of nucleoid proteins on DNA repression loop formation in Escherichia coli. **35,** 3988–4000 (2007).
82. Lioy, V. S. *et al.* Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell* **172,** 771–783.e18 (2018).
83. Schatz, D. G. & Ji, Y. Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* **11,** 251–263 (2011).
84. Kar, S., Edgar, R. & Adhya, S. Nucleoid remodeling by an altered HU protein: reorganization of the transcription program. *Proc Natl Acad Sci USA* **102,** 16397–16402 (2005).
85. Balandina, A., Kamashev, D. & Rouviere-Yaniv, J. The bacterial histone-like protein HU specifically recognizes similar structures in all nucleic acids. DNA, RNA, and their hybrids. *J. Biol. Chem.* **277,** 27622–27628 (2002).
86. Macvanin, M. *et al.* Noncoding RNAs binding to the nucleoid protein HU in Escherichia coli. *Journal of Bacteriology* **194,** 6046–6055 (2012).
87. Qian, Z. *et al.* A New Noncoding RNA Arranges Bacterial Chromosome Organization. *MBio* **6,** e00998–15 (2015).
88. Qian, Z., Zhurkin, V. B. & Adhya, S. DNA-RNA interactions are critical for chromosome condensation in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **114,** 12225–12230 (2017).
89. Dame, R. T., Hall, M. A. & Wang, M. D. Single-molecule unzipping force analysis of HU-DNA complexes. *Chembiochem* **14,** 1954–1957 (2013).
90. Hammel, M. *et al.* HU multimerization shift controls nucleoid compaction. *Sci Adv* **2,** e1600650 (2016).
91. Wang, S., Moffitt, J. R., Dempsey, G. T., Xie, X. S. & Zhuang, X. Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 8452–8457 (2014).
92. Lal, A. *et al.* Genome scale patterns of supercoiling in a bacterial chromosome. *Nat Commun* **7,** 11055 (2016).
93. Javer, A. *et al.* Short-time movement of E. coli chromosomal loci depends on coordinate and subcellular localization. *Nat Commun* **4,** 3003 (2013).
94. Lau, I. F. *et al.* Spatial and temporal organization of replicating Escherichia coli chromosomes. *Mol. Microbiol.* **49,** 731–743 (2003).

95. Hensel, Z., Weng, X., Lagda, A. C. & Xiao, J. Transcription-factor-mediated DNA looping probed by high-resolution, single-molecule imaging in live E. coli cells. *PLoS Biol.* **11,** e1001591 (2013).

96. Jen-Jacobson, L., Engler, L. E. & Jacobson, L. A. Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure* **8,** 1015–1023 (2000).

97. Aki, T., Choy, H. E. & Adhya, S. Histone-like protein HU as a specific transcriptional regulator: co-factor role in repression of gal transcription by GAL repressor. *Genes to Cells* **1,** 179–188 (1996).

98. Aki, T. & Adhya, S. Repressor induced site-specific binding of HU for transcriptional regulation. *The EMBO Journal* **16,** 3666–3674 (1997).

99. Swinger, K. K., Lemberg, K. M., Zhang, Y. & Rice, P. A. Flexible DNA bending in HU-DNA cocrystal structures. *The EMBO Journal* **22,** 3749–3760 (2003).

100. Dorman, C. J. Function of nucleoid-associated proteins in chromosome structuring and transcriptional regulation. *J. Mol. Microbiol. Biotechnol.* **24,** 316–331 (2014).

101. Schlünzen, F. *et al.* Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **413,** 814–821 (2001).

102. Mosteller, R. D. & Yanofsky, C. Transcription of the tryptophan operon in Escherichia coli: rifampicin as an inhibitor of initiation. *J. Mol. Biol.* **48,** 525–531 (1970).

103. Chen, H., Shiroguchi, K., Ge, H. & Xie, X. S. Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli. *Molecular Systems Biology* **11,** 781–781 (2015).

104. Gabain, von, A., Belasco, J. G., Schottel, J. L., Chang, A. C. & Cohen, S. N. Decay of mRNA in Escherichia coli: investigation of the fate of specific segments of transcripts. *Proc Natl Acad Sci USA* **80,** 653–657 (1983).

105. Bakshi, S., Siryaporn, A., Goulian, M. & Weisshaar, J. C. Superresolution imaging of ribosomes and RNA polymerase in live Escherichia coli cells. *Mol. Microbiol.* **85,** 21–38 (2012).

106. Bakshi, S., Choi, H., Mondal, J. & Weisshaar, J. C. Time-dependent effects of transcription- and translation-halting drugs on the spatial distributions of the Escherichia coli chromosome and ribosomes. *Mol. Microbiol.* **94,** 871–887 (2014).

107. Wada, M., Kano, Y., Ogawa, T., Okazaki, T. & Imamoto, F. Construction and characterization of the deletion mutant of hupA and hupB genes in Escherichia coli. *J. Mol. Biol.* **204,** 581–591 (1988).

108. Fei, J. *et al.* RNA biochemistry. Determination of in vivo target search kinetics of regulatory noncoding RNA. *Science* **347,** 1371–1374 (2015).
109. Dayton, C. J., Prosen, D. E., Parker, K. L. & Cech, C. L. Kinetic measurements of Escherichia coli RNA polymerase association with bacteriophage T7 early promoters. *J. Biol. Chem.* **259,** 1616–1621 (1984).
110. Agresti, A., Scaffidi, P., Riva, A., Caiolfa, V. R. & Bianchi, M. E. GR and HMGB1 interact only within chromatin and influence each other's residence time. *Molecular Cell* **18,** 109–121 (2005).
111. Pallier, C. *et al.* Association of chromatin proteins high mobility group box (HMGB) 1 and HMGB2 with mitotic chromosomes. *Mol. Biol. Cell* **14,** 3414–3426 (2003).
112. Brázda, V., Laister, R. C., Jagelská, E. B. & Arrowsmith, C. Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol. Biol.* **12,** 33 (2011).
113. Schavemaker, P. E., Śmigiel, W. M. & Poolman, B. Ribosome surface properties may impose limits on the nature of the cytoplasmic proteome. *eLife Sciences* **6,** 21 (2017).
114. Bohrer, C. H., Bettridge, K. & Xiao, J. Reduction of Confinement Error in Single-Molecule Tracking in Live Bacterial Cells Using SPICER. *Biophysical Journal* **112,** 568–574 (2017).
115. Kusumi, A., Sako, Y. & Yamamoto, M. Confined lateral diffusion of membrane receptors as studied by single particle tracking (nanovid microscopy). Effects of calcium-induced differentiation in cultured epithelial cells. *Biophysical Journal* **65,** 2021–2040 (1993).
116. Darzacq, X. *et al.* In vivo dynamics of RNA polymerase II transcription. *Nature Structural & Molecular Biology* **14,** 796–806 (2007).
117. Weng, X., Bohrer C.H., Bettridge, K., Lagda, A.C., Cagliero, C., Jin, D. J., & Xiao, J. RNA polymerase organizes into distinct spatial clusters independent of ribosomal RNA transcription in *E. coli*. *bioRxiv* 320481 (2018).
118. Abad, E., Yuste, S. B. & Lindenberg, K. Reaction-subdiffusion and reaction-superdiffusion equations for evanescent particles performing continuous-time random walks. *Phys. Rev. E* **81,** 031115 (2010).
119. Gahlmann, A. & Moerner, W. E. Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging. *Nature Reviews Microbiology* **12,** 9–22 (2014).

CURRICULUM VITAE

Kelsey Bettridge
kelsey.bettridge@gmail.com
Place of birth: Lafayette, Indiana
Date last updated: March 22, 2019

EDUCATION

Johns Hopkins University, School of Medicine Baltimore, MD
Doctor of Philosophy in Biophysics April 2019

Lawrence University Appleton, WI
Bachelor of Arts in Physics, graduated *cum laude* June 2011

RESEARCH EXPERIENCE

Johns Hopkins University, School of Medicine
Baltimore, MD
Graduate Student Researcher…….…………………………...June 2013 – present
Lab of Dr. Jie Xiao, Department of Biophysics and Biophysical Chemistry

- Investigated the dynamics of RNA polymerase using fluorescence recovery after photobleaching (FRAP), super-resolution single molecule imaging, image processing, and advanced statistical analysis using MATLAB and JavaScript programming languages
- Collaborated with the lab of Dr. Young at Texas A&M on a project in which we explored the *in vivo* structure of a bacteriophage lysis protein, holin, in *E. coli* using super-resolution single molecule imaging
- Investigated the roles of the various binding modes of the nucleoid-associated protein HU in live *E. coli* cells using superresolution single molecule tracking along with genetic manipulations.

Research Technologist…………………………..……….June 2011 – August 2012
Lab of Dr. Carol Greider, Department of Molecular Biology and Genetics

- Researched the role of a telomere-associated protein, Rif2, in its ability to help telomerase preferentially elongate short telomeres through systematic protein mutagenesis and molecular biology techniques, such as Southern and Western blots

PUBLICATIONS

- Xiaoli Weng*, Chris Bohrer*, **Kelsey Bettridge**, Arvin Ladga, Din Jin, and Jie Xiao. “RNA polymerase organizes into distinct spatial clusters independent of ribosomal RNA transcription in *E. coli*.” *Proceedings of the National Academies of Science* (in submission).
- **Bettridge, Kelsey**, Ashley Cook, Roy Ziegelstein, and Peter Espenshade. “A Scientist’s Oath.” *Molecular Cell* 71.6 (2018): 879-881.
- Bohrer, Christopher H., **Kelsey Bettridge**, and Jie Xiao. "Reduction of Confinement Error in Single-Molecule Tracking in Live Bacterial Cells Using SPICER." *Biophysical Journal* 112.4 (2017): 568-574.
- Kaizer, Hannah, et al. "Regulation of Telomere Length Requires a Conserved N-Terminal Domain of Rif2 in *Saccharomyces cerevisiae*." *Genetics* 201.2 (2015): 573-586.

PRESENTATIONS AND ABSTRACTS

- **Bettridge, Kelsey**, Xiaoli Weng, Subhash Verma, Sankar Adhya, and Jie Xiao. “Probing the role of DNA-binding modes on the dynamics of HU in live *E. coli* cells”, oral presentation, Biophysical Society Annual Meeting, Baltimore, MD, March 5th, 2019.
- **Bettridge, Kelsey**, Xiaoli Weng, Subhash Verma, Sankar Adhya, and Jie Xiao. “Nucleoid-associated protein HU dynamics depend on REP325 ncRNAs”, oral presentation, Hopkins RNA Salon, Baltimore MD, November 30, 2018
- **Bettridge, Kelsey**, Xiaoli Weng, Subhash Verma, Sankar Adhya, and Jie Xiao. “Nucleoid-associated protein HU dynamics depend on REP325 ncRNAs”, oral presentation, 4th Chesapeake Bay Area Single Molecule Meeting, November 10, 2018
- **Bettridge, Kelsey**, Chris Bohrer, and Jie Xiao. “Probing the search dynamics of RNA polymerase in live E. coli cells”, published abstract, Single Biomolecules Meeting at Cold Spring Harbor, NY, August 29, 2018
- **Bettridge, Kelsey,** Chris Bohrer, and Jie Xiao. “Investing the promoter search dynamics of RNA polymerase in live *E. coli* cells: a ‘Where’s Waldo?’ story”, oral presentation, ASCB conference, Philadelphia, PA, December 5, 2017
- **Bettridge, Kelsey**, Chris Bohrer, and Jie Xiao. “Probing the search dynamics of RNA polymerase in live *E. coli* cells using single molecule and

statistical methods", oral presentation, Biophysical Society Annual Conference, New Orleans, LA, February 14, 2017

LEADERSHIP, MENTORING, AND VOLUNTEER EXPERIENCE

Johns Hopkins University, School of Medicine
Baltimore, MD

Graduate Student Association roles

President ……………………………………..…………..…….July 2016 – June 2017

- Managed an executive board of seven members and supervised budget of $32,000
- Wrote three new policies and revised the GSA constitution to further improve transparency, clarify procedures, and improve efficiency
- Improved relationships with other student-serving administrative offices, and collaborated with three offices to complete nine programming events

President-Elect ………………………………………….……....July 2015 – June 2016

- Clarified and revised two existing policies to improve transparency and efficiency
- Served on two upper administrative committees, and supervised a student committee dedicated to fundraising for student conference travel awards

Program Representative…………………………..…….October 2014 – June 2015

- Served as the graduate student representative on the Faculty Senate committee
- Organized five, 4-hour volunteer experiences in collaboration with local community service partner Civic Works for graduate students

Biophysics for Baltimore Teens and Student Academic Research Experience programs

Mentor/Teacher…………………………..………………Summer 2015 and 2016

- Mentored two high school students, training them to do molecular biology, imaging techniques, and basic programming
- Coached the two students in professionalism and science communication
- For BRBT, prepared two lectures for the students on various basic scientific topics

TEACHING EXPERIENCE

Johns Hopkins University, School of Medicine
Baltimore, MD

Teaching Assistant …………………………………………….……….Spring 2016

- Lead weekly group tutoring sessions to provide assistance for homework and conceptual problems
- Graded ten sets of homework and three tests for a class of forty participants

Self-Employed
Baltimore, MD
Tutor for introductory science courses……….……………….January 2012 – 2017

- Tutored eight students individually on various introductory-level science courses, some students for several years
- Developed custom-tailored tutoring and lesson plans for each student

Lawrence University
Appleton, WI
Tutor for astronomy and classical physics……….……January 2009 – June 2011

- Lead weekly group and individual tutoring sessions for astronomy and classical mechanics course participants
- Graded eight weekly homework sets for a class of over fifty people for three semesters

HONORS AND AWARDS

- Selected to present work and co-chair a session at the Annual Biophysical Society Meeting in March 2019
- Carlson Award for Excellence in Biophysics, 2012 academic year
- Initiated into Sigma Pi Sigma physics honor society, 2010